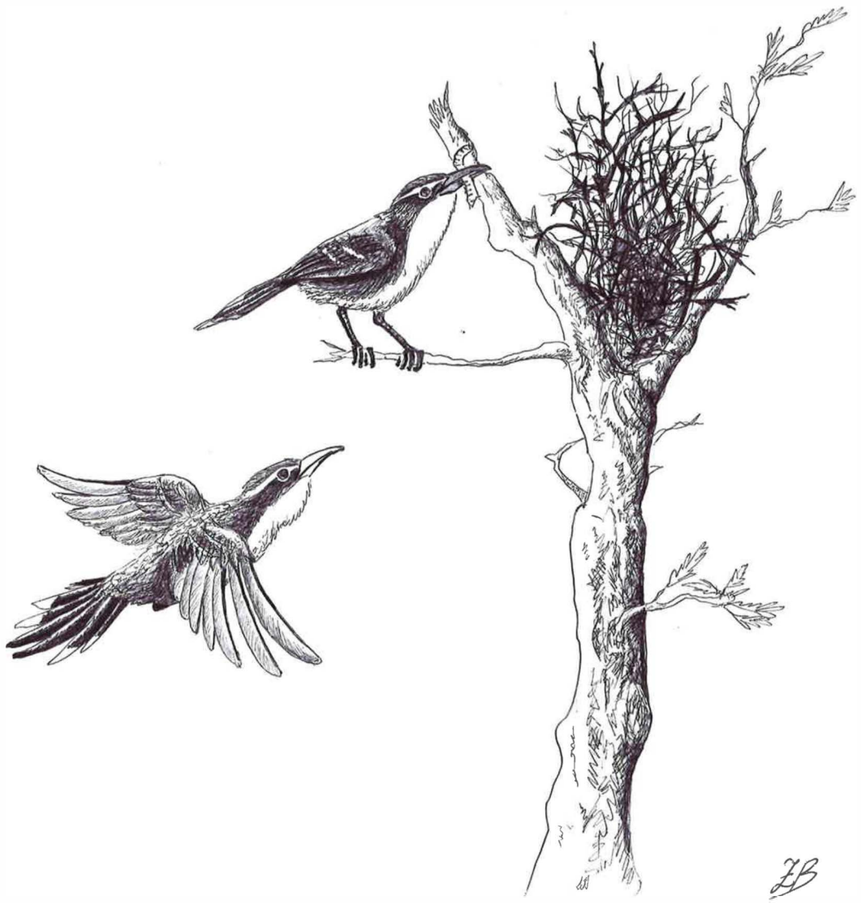


Dissertation zur Erlangung der naturwissenschaftlichen Doktorwürde (Dr. sc. nat.)
vorgelegt der Mathematisch-naturwissenschaftlichen Fakultät der Universität Zürich

Vocal Combinations in the Southern Pied Babbler (*Turdoides bicolor*) and the Chestnut-Crowned Babbler (*Pomatostomus ruficeps*): Implications for the Evolution of Human Language

von

Sabrina Engesser

aus Deutschland

Zürich, 2017

Promotionskommission:
Dr. Simon William Townsend (Leitung der Dissertation)
Prof. Dr. Marta Berta Manser (Vorsitz und Leitung der Dissertation)
Dr. Amanda Ruth Ridley
Prof. Dr. Carolus van Schaik

Vocal Combinations in the Southern Pied Babbler (*Turdoides bicolor*) and the Chestnut-Crowned Babbler (*Pomatostomus ruficeps*): Implications for the Evolution of Human Language

---

DISSERTATION

ZUR

ERLANGUNG DER NATURWISSENSCHAFTLICHEN DOKTORWÜRDE

(DR. SC. NAT.)

VORGELEGT DER

MATHEMATISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT

DER

UNIVERSITÄT ZÜRICH

VON

SABRINA ENGESSER

AUS

DEUTSCHLAND

PROMOTIONSKOMMISSION

DR. SIMON WILLIAM TOWNSEND (LEITUNG DER DISSERTATION)

PROF. DR. MARTA BERTA MANSER (VORSITZ UND LEITUNG DER DISSERTATION)

DR. AMANDA RUTH RIDLEY

PROF. DR. CAROLUS VAN SCHAIK

ZÜRICH, 2017

Cover illustrations by Zinaida Bogdanova

CONTENTS

SUMMARY

Language's expressive power is one of its key characterising features. This generative capacity is achieved through language's double articulatory nature: meaningless sounds (phonemes) are combined to create meaningful words (phonology/combinatoriality), and words are assembled into higher-order meaningful phrases (syntax/compositionality). Comparative work on non-human animals investigating the evolutionary origin of combinatorial abilities has so far focused on singing species or on primates. Although these studies have shed light on the combinatorial capacities outside of humans, evidence for basic phoneme-like or semantically compositional structures in non-human communication systems is rare. By taking a comparative approach, investigating the prevalence and diversity of combinatoriality within the discrete call system of two highly social passerine birds, this dissertation aimed to unveil selective drivers promoting combinatorial capacities, and provides analogue examples to, and potential precursors of, language's combinatorial layers.

Work on chestnut-crowned babblers (*Pomatostomus ruficeps*) demonstrates the reuse of two meaningless sounds (*A* & *B*) in different arrangements to generate the functionally distinct *AB*/'flight' call (a contact call) and *BAB*/'prompt' call (a food-provisioning call). Specifically, the meaning differentiation between the two calls was found to be the result of a single modification, akin to a rudimentary phonemic-like contrast in human language. Work on free-living, habituated southern pied babblers (*Turdoides bicolor*) revealed that upon encountering predators the birds combine 'alert' with 'recruitment' calls into a 'mobbing sequence'. While alert calls were found to be produced in response to low-urgency threats and recruitment calls during group travel, the mobbing sequence seems to combine information on the threat and the requested action with receivers parsing the sequence in a compositional way. Further investigations into the pied babbler recruitment call component indicated that recruitment calls can additionally be produced in form of two longer 'recruitment cry' variants, composed of either repetitions of *A'*/single-note motifs or *A'B'*/double-note motifs respectively. Although both cries were found to function in recruiting the group during group travel, the cries' internal structure seemed to specify the form of recruitment, either requesting approach or prompting following over long distances. Accordingly, recruitment cries appear to be characterised by meaning-differentiating variation at the internal structure level of a larger meaningful signal. Lastly, this dissertation provides evidence for meaningful temporal structuring within pied babbler 'clucks' and 'purrs'. While both calls were shown to be composed of repetitions of the same acoustic element, the number of element repetitions was found to encode qualitatively different information. Specifically, longer purrs function to attract dependent offspring to a food source and truncated clucks seem to communally mediate imminent foraging site switches.

The empirical data provided in this dissertation demonstrates that babblers utilise various combinatorial mechanisms, and furthermore indicates that neither meaningful combinations generated from acoustically differentiable meaningless elements, or rudimentary compositional structures, are unique to human language. By demonstrating evidence for meaningful vocal combinations in species distantly related to humans, this work provides fundamental insights into the factors that might have promoted the evolutionary progression of language's generative system.

ZUSAMMENFASSUNG

Die sprachliche Kreativität ist eine der Haupteigenschaften der menschlichen Sprache. Diese Eigenschaft aus begrenzten Lauten unendlich Gebrauch zu machen, ist das Resultat kombinatorischer Mechanismen auf zweier Ebenen (zweifache Gliederung): Bedeutungslose Laute (Phoneme) werden zu bedeutungsvollen Wörtern zusammengesetzt (Phonologie/ Kombinatorialität), und Wörter wiederum zu höher geordneten Phrasen oder Sätzen (Syntax/ Kompositionalität). Vergleichende Studien an Tieren, mit dem Ziel den evolutionären Ursprung solch kombinatorischer Fähigkeiten zu untersuchen, haben sich bislang singenden Arten oder Primaten gewidmet. Während jene Studien wichtige Einblicke über kombinatorische Fähigkeiten in tierischen Kommunikationssystemen erbrachten, bleibt unklar, ob fundamentale phonemische und kompositionale Strukturen ebenfalls in der Kommunikation von Tieren vertreten sind. Diese Dissertation hat zum Ziel kombinatorische Mechanismen in den diskreten Rufen zweier hoch sozialer, in Gruppen lebender Vögel zu untersuchen: den Elsterdrosslingen (*Turdoides bicolor*) und den Rotscheitelsäblern (*Pomatostomus ruficeps*). Solch eine vergleichende Studie in zwei zum Menschen entfernt verwandten Tierarten kann helfen, selektive Faktoren und evolutionäre Vorstufen in der Entstehungsgeschichte der sprachlichen Kreativität und ihrer kombinatorischen Stufen (Phonologie & Syntax) aufzudecken.

Zum Ersten zeigt diese Dissertation, dass wilde Rotscheitelsäbler zwei bedeutungslose akustische Elemente (*A* & *B*) in unterschiedlichen Konstellationen wiederverwenden um zwei bedeutungsvolle, in ihrer Funktion unterschiedliche, Rufe zu generieren: den *AB*-Flugruf und den *BAB*-Fütterungsruf. Experimente demonstrierten, dass der Bedeutungsunterschied zwischen beiden Rufen durch die erste *B*-Note des Fütterungsrufs generiert wird, und somit einen rudimentären Phonem-Kontrast repräsentiert. Die Arbeit an Elsterdrosslingen zeigt auf, dass diese in Gegenwart von Fressfeinden Alarmrufe und Rekrutierungsrufe zu einer Art Mobbing-Sequenz kombinieren. Während Alarmrufe als Reaktion auf plötzliche, aber generell wenig gefährliche, Gefahren erfolgten, und Rekrutierungsrufe während Streifzügen durch das Territorium, scheint die Mobbing-Sequenz die Bedeutung beider Rufe zu vereinen und dementsprechend Information über die Gefahr und die angeforderte Aktion zu kombinieren. Demzufolge stellt die Sequenz eine elementare Komposition dar, deren übergeordnete Bedeutung sich aus der Bedeutung ihrer individuellen Teile ableitet. Des Weiteren wird gezeigt, dass der Rekrutierungsruf in Form zweier längerer Rekrutierungs-Sequenzen wiedergegeben werden kann, wobei die Sequenzen aus sich wiederholenden Motiven zusammengesetzt sind: einsilbigen *A'*-Motiven oder doppelsilbigen *A'B'*-Motiven. Während beide Rekrutierungs-Sequenzen gleichermassen dazu dienen die Gruppenmitglieder zu rekrutieren, spezifiziert die innere Struktur die Art der Rekrutierung. Demnach führen Sequenzen aus einsilbigen Motiven zur Annäherung der Gruppenmitglieder an den Signalgeber, wohingegen Sequenzen aus doppelsilbigen Motiven die Gruppe animieren dem Signalgeber über weite Strecken zu folgen. Zu guter Letzt zeigt diese Arbeit, dass Elsterdrosslinge zwei unterschiedlich pulsierende Rufe generieren, welche aus dem identischen akustischen Element bestehen und sich lediglich in der Anzahl wiederholender Elemente unterscheiden. Zum einen produzieren Elsterdrosslinge sogenannte 'purrs' (Schnurrlaute), welche aus circa 14 Element-Repetitionen bestehen, und welche dazu dienen Jungvögel zu Futterplätzen zu locken. Zum anderen erzeugen sie die aus nur 2-3 Elementen

bestehenden ‘clucks’ (Gurrlaute), welche gemeinschaftlich von der ganzen Gruppe produziert werden um bevorstehende Ortswechsel zu vermitteln.

Zusammenfassend demonstriert diese Dissertation, dass Rotscheitelsäbler und Elsterdrosslinge diverse kombinatorische Mechanismen verwenden, und liefert die Ersten empirischen Beweise für rudimentäre phonemische und kompositionale Strukturen ausserhalb der menschlichen Sprache. Der Nachweis solcher Strukturen in vom Menschen entfernt verwandten Tierarten gibt fundamentale Einblicke darüber, welche Faktoren die Entstehung der sprachlichen Generativität vorangetrieben haben könnten.

ACKNOWLEDGEMENTS

Firstly, I thank Simon Townsend for his supervision on this dissertation. I am deeply grateful for his guidance and his faith in me, and the way he introduced me into the world of science. I also thank Amanda Ridley for permitting work on her charming pied babblers, and for her easy-going and uncomplicated encouragement and support. I am much obliged to her for giving me the opportunity to work on such lovely and well-habituated birds, and for letting me be a part of the team babbler. Although they are far less fluffy and friendly than the pieds, I am also thankful to Andrew Russell for letting me collaborate on the handsome, but wild-wild chestnut-crowned babblers, and for the craziness and fun in the field. I am indebted to Marta Manser for her support throughout my dissertation and for the inspiring discussions. I also thank Tim Clutton-Brock for permission to work at the Kuruman River Reserve. I am very grateful to Marianne Köpfler; her commitment and friendly ear often helped me let off steam and see things from different perspectives. Thanks also to Isabel Schöchli for being a very warm-hearted person who never hesitates to help.

I met wonderful people over the last years: Andrea Liebl, Andri Manser, Ariana Stransburg-Peshkin, Bart Kranstauber, Bruce Boatman, Christophe Bousquet, Debbie Leigh, Elliot Capp, Eva Cereghetti, Eva Koch, James Westrip, Jan Runge, Jari Garbely, Jobran Chebib, Margot Oorebeek, Marljin Henskens, Matt Creasey, Max Schmid, Megan Wyman, Naomi Zweerus, Nina Vasiljevic, Patricia Lopes, Paul Haverkamp, Pauline Toni, Pheadra Kokkini, Sally Steinert, Sofia Grize, Tejasvini Chalikonda, Tim Vink & Yannick Auclair; some of which became good friends and who made times colourful and lively: Denise Karp, Gabriella Gall, Ramona Rauber, Inês Gonçalves, Jamie Samson, Jelena Mausbach, Nicola Harrison, Robbi Hopper & Sandra Röthlisberger. Extra special thanks go to Elizabeth Wiley, Marianna Perez, Ruan De Bruin, Rute Mendonça & Tanja Van de Ven who no matter how many weeks or months lie between chats, feel like true companions, never really separated in time or space.

Special thanks go to Elodie Mandel-Briefer for always being supportive when I desperately needed advice on acoustics; Zinaida Bogdanova for the amazing babbler drawings; and silly-billy Andreas Sutter who gave me some good laughs and helped me a few times with stats when I felt like smashing the computer. Big thanks also to Katie Collier for being an absolutely joyful and fun officemate, for being my living dictionary and grammar checker, and for the great and inspiring discussions we had. I am particularly grateful to have gotten to know Manuela Ferrari, who (already at the time when we were still Master students) became a close companion and a valuable friend, always encouraging me, no matter whether it was in terms of (teaching me!) stats, discussing my work, or motivating me when things seemed dark and grey.

Throughout the last 11 years, the most supportive person has always been my partner Sebastian Seel. Although never having worked in the scientific field, he has always been one of the most understanding persons, truly grasping the work I am doing and always backing me up with words and deeds. I am also thankful for my family, who (even if not fully understanding what exactly it is that I am doing) always showed great interest in my work and helped me whenever I needed them.

This dissertation was funded by a 16-month and 10-month studentship allocated by the Forschungskredit of the University of Zurich (grant no. 57191601 & FK-14-077) and by a 3 month Doc.mobility fellowship granted by the Swiss National Science Foundation (grant no. P1ZHP3_151648); all the grants being awarded to myself under the supervision of Dr. Simon William Townsend. A further 16-month grant of the Swiss National Science Foundation (31003_153065) was awarded to Dr. Simon William Townsend for this dissertation.

CHAPTER 1

General introduction

Vocal combinations in non-human animals

Research over the last five decades has indicated that numerous aspects of human language also exist in non-human communication systems [1, 2]. Reference and intentionality represent two key components of language, with meaning being assigned to vocal structures, and information being voluntarily communicated [1]. Analogue forms of these components are found in various forms in non-human species. Animal vocalisations can, for example, refer to current external events or objects [3, 4], and signals can be flexibly used by animals to inform or manipulate receivers, or equally, information can be withheld in the presence or absence of certain individuals [5, 6]. Such strategic, flexible use of vocalisations indicates that vocalisations and the decision to call are not necessarily hardwired in animals, but individuals might have a certain degree of control over their vocal production [7]. Whilst these findings have been argued to provide insights into understanding the evolution of linguistic abilities central to language, there remains a problem with regard to language's generative nature, particularly its evolutionary origin and the selective conditions promoting its emergence [2, 8, 9].

Linguistic generativity facilitates the expression of limitless thoughts and ideas, and is, for the most part, achieved as a result of its double articulation [1, 10-13]. At the phonological level, human language expressions are considered combinatorial [1]. Phonemes such as /ɪ/, /n/, and /t/, themselves are meaningless, but when produced together in a specific order can generate meaningful morphemes or words such as 'in-' [ɪn], 'tin' [tɪn], and 'nit' [nɪt] [8]. At the higher syntactic layer, these constructs can then themselves be combined together to produce semantically compositional phrases or sentences, where the meaning of the composition is derived from the independent meaning of the individual units, and the set of rules governing their sequential organisation [14, 15]. Theoretical work hypothesises that language's combinatorial layers evolved in order to overcome productional and perceptional limitations [16]. Specifically, stringing meaningless sounds (phonemes) together can enhance the discriminability between otherwise similar sounding signals, and hence decrease perception mistakes [16, 17]. Once the number of messages to be encoded exceeds the number of discrete signals present in a communicative system, and in order to offset memory limitations, meaningful signals can then be assembled in a systematic way into higher order meaningful structures [16, 18, 19].

Empirical data on animal communication systems can help to test such hypotheses, and a broad comparative approach can provide insights into the evolutionary progression of human language's combinatorial components [2]. Historically, comparative work investigating combinatorial abilities in non-human animals has focused on primate species [2]. In fact, studies on guenon monkeys (*Cercopithecus* spp.) revealed some parallels to language's syntactic layer, with monkeys combining meaningful calls into higher-order meaningful structures [20, 21]. Although such empirical data from primates can unveil potential precursors and shed light on the phylogenetic origins of language specific features, work on more distantly related species are key in order to elucidate the environmental and social factors promoting the evolution of combinatorial mechanisms [2]. Accordingly, such analogous data in non-primate species can have important implications with regard to

convergent evolutionary mechanisms [2]. Studies investigating phonological capacities have provided such an analogy between the sequential organisations of meaningless sounds in animal songs and words in language [22, 23]. However, while songs are structurally complex, in contrast to language's phonological system, they lack semantics, and changes within a song's structure are largely irrelevant for the encoded type of message and do not differentiate meaning [22, 23]. Although lacking basic components of a phonological system, learning plays a central role in the acquisition of songs in animals and hence such systems are still considered key candidates to study the neural and developmental mechanisms of language acquisition [2, 22, 23].

In line with the comparative approach, with this dissertation I aimed to investigate the combinatorial abilities of two social passerine bird species that, unlike many songbirds, do not sing, but instead produce an array of discrete, meaningful vocalisations [24, 25]. By investigating species distantly related to humans, but with mammal-like vocalisations, the evolutionary factors driving the emergence of both phonology and syntax may be more adequately addressed. The following sections will briefly review the forms of combinatorial mechanisms so far described in non-human animals, also considering examples that do not provide an apparent or adequate analogue to language's levels of articulation (see Table 1 for summary). Sounds will be referred to as meaning*less* if they represent "acoustically differentiable elements" [26] that are not emitted in isolation, and hence are unlikely to transfer functionally relevant information. Acoustic units are considered meaning*ful* if their production is context-specific and elicit predictable responses in receivers, suggesting they serve a distinct function and transfer qualitatively different information.

Types of combinatorial mechanisms

Phonocoding (songs): A range of animals, including songbirds (Passeriformes), gibbons (Hylobatidae), hyraxes (Procaviidae) and whales (Cetartiodactyla), combine meaningless elements into higher-order, often hierarchically structured, sequences or songs [27-31]. Such combinatorial structures predominantly function in territorial or courtship display [28, 29, 32], to facilitate recognition between individuals or groups [28, 33-40], or to strengthen the bonding among partners or groups [40, 41]. Although on the surface level such sequences or songs resemble language's phonological layer, they lack fundamental features of a basic phonological system, not carrying any propositional meaning, and element arrangement being insignificant for the encoded type of information [22, 26, 28]. While words in language are composed of phonemes, the elements of songs are more accurately defined as phones, since they lack meaning-differentiating characteristics [13, 32]. In sum, animal songs are typically considered to represent a form of phonocoding [42, 43].

Behaviour specific combinatoriality: Chickadees, tits and titmice (Paridae) are known to generate dozens of combinatorial call-variants by omitting or duplicating distinct (potentially meaningless) sound elements within a fixed-ordered sequence of the call [44-47]. Although there seems to be a degree of context specificity with some call variants being more likely produced during certain behaviours (potentially linked to locomotion

and movement [44]) [46-48], and/or changes in the structure modify receiver responses during playback experiments [49, 50], the information encoded in a note composition remains largely ambiguous [44, 46, 51]. Accordingly, combinations of meaningless sound elements that generate a call (or multiple calls) produced during particular behaviours may represent a form of 'behaviour specific combinatoriality', distinguishable from 'functionally distinct combinatoriality' where combinatorial variation may transfer functionally more relevant, qualitatively distinct and meaningful information.

Functionally distinct combinatoriality: Examples, where the rearrangement of meaningless sounds transfers functionally distinct, meaningful information for receivers are, to date, absent. Meaning modifications could be achieved through deleting, duplicating or exchanging (single) elements within a sequence of meaningless sounds (akin to the parid call system [44, 46, 47]). Such meaningless, yet meaning-differentiating elements may represent an analogue to phonemes or phoneme-like structuring in human language.

Mixed compound calls: Wedge-capped capuchins (*Cebus olivaceus*) produce compound calls of two, three or four different call types [52]. While a certain call type is associated with a particular emotional or motivational state, compound calls appear to be produced in situation where individuals experience conflicting interests or motivations (e.g. submission vs. aggression) [52]. Accordingly, such compound calls appear to reflect intermediate states of the caller, not encoding information compositionally [53].

Semantically combinatorial structures: Meaningful units or calls can further be assembled into sequences that encode new information, that is by definition unrelated to the meaning of its component calls [54]. For example, putty-nosed monkeys (*Cercopithecus nictitans*) combine predator specific alarm calls into longer sequences that elicit group movement in non-predatory context [21, 55, 56]. Gorillas (*Gorilla beringei beringei* & *G. gorilla*) produce context specific vocalisations during social interactions, which can be assembled into larger sequences that appear to share no or little overlapping context with the individual contexts the constituents are produced in [57]. Similarly, Campbell's monkeys (*Cercopithecus campbelli campbelli*) and gibbons produce context specific, structurally unique sequences in predatory and social contexts, by recombining a fixed set of meaningful and meaningless (i.e. not produced in isolation) call types [20, 58]. Although the parts, or some parts, of the sequences are presumed to carry semantic information, the structures themselves do not appear to be processed in a systematic, compositional way, since the meaning of the whole cannot be derived from its compounds. As such the sequences are suggested to constitute semantically combinatorial structures [8, 54].

Segmental concatenation: A few mammalian species have been described to concatenate acoustic segments in a more systematic way. For example, Diana monkeys (*Cercopithecus diana*), banded mongooses (*Mungos mungo*) and dingoes (*Canis lupus dingo*) produce an individually distinct segment that can be produced in isolation or concatenated with other distinct or graded elements, correlating with the animal's behaviour or motivation [59-62]. From a linguistic perspective, such segments may resemble morphemes (smallest meaningful units), with the individually distinct elements representing free morphemes

that can be produced as a standalone segment, or be conjoined with the behaviour- or motivation-coding segment representing bound morphemes [13, 43].

Affixation: Similar to segmental concatenations, Campbell's monkeys produce two predator specific alarm calls, which can be produced in isolation or concatenated with a meaning-modifying acoustic element [63, 64]. The modifier functions to broaden the meaning of the predator specific alarm calls in a predictable way, and converts them into general disturbance calls [63, 64]. However, in the previous examples of segmental concatenations, the standalone (individuality-coding) segment was fixed and the bound segments (behaviour- or motivation-coding) were variable. The opposite is true for Campbell's monkeys, with the predator specific alarm calls being variable standalone segments, and the bound modifier being fixed, akin to an affix [63-65]. In linguistic terms, the affixation rule applied by Campbell's monkeys can be considered a (rudimentary) sub-form of compositional syntax, with the affix modifying either alarm call in a conserved way, and hence the combinations' meaning reflecting the meanings of its parts [43].

Semantically compositional structures: The closest example to compositional syntax outside of humans so far described is found in the vocal system of Japanese great tits (*Parus major minor*) [66]. Specifically, the birds combine warning and recruitment calls. While the first elicits a vigilance response and the latter an approach to the caller, the combination elicits a mix of both behaviours [66]. Although the study posits that the two vocalisations encode different meaning with the combination encoding a deduced compound meaning [66], the construct might alternatively represent an 'mixed compound call', since the behavioural response to the combination appears to be intermediate to the responses elicited by the individual calls (see also general discussion) [66].

Temporality: Besides the combination of acoustic elements or units, new information can additionally be encoded through varying the temporal arrangement within a sequence of repeated elements. The most commonly described functions of temporal modifications is to transfer information on an individual's arousal state experienced when encountering conspecifics or predators [67-76]. In this case, information is generally encoded by gradual changes in the number or the rate of repeated elements or inter-element intervals [67-76]. While such temporal gradients generally correlate with motivational or urgency levels experienced by the caller, work on the alarm call system of colobus monkeys (*Colobus guereza* & *C. polykomos*) and the social calls of Mexican free-tailed bats (*Tadarida brasiliensis*) has shown temporal structures can further encode categorical information, distinguishing predator types or behavioural contexts for example [77-79].

Table 1. Summary of terms and definitions of types of combinatorial structures in animal vocal systems, including examples from the English language, occurrences and functions. Grey annotations represent clarifications, with small letters denoting meaningless sounds and capital letters context specific or meaningful units.

Type	Definition	Examples	Occurrences	Functions
Phonocoding	Combination of meaningless sounds into sequences without propositional meaning a b c …	-	Songs of birds, gibbons, hyraxes, whales; dolphin whistles [27-41]	Territorial/courtship display [28-30, 32]; Individual/group recognition [28, 33-40]; Bonding behaviour [40, 41]
Behaviour specific combinatoriality	Combination of meaningless elements to create behaviour specific calls a b c d = behaviour X a c d d = behaviour Y	-	Calls of chickadees, tits, titmice [44-47, 49-51]	Potentially locomotion or movement specific [44]
Functionally distinct combinatoriality	Combination of meaningless elements to create functionally distinct meaningful calls a b c = meaning X b a c = meaning Y	*cat* *act*	-	-
Mixed compound calls	Combination of motivation/emotion-coding calls reflecting intermediate states of the caller X = motivation/emotion X Y = motivation/emotion Y X Y = motivation/emotion < X + Y	Mixture of aggressive and submissive behaviour [53]	Calls of wedged-capped capuchins [52]	Affiliative and agonistic social contexts [52]
Semantically combinatorial structures	Combination of calls with meaning unrelated to the meaning of its constituents X = meaning X Y = Y (meaningful or meaningless) X + Y = meaning Z	*hot* *potato* *hot potato*	Putty-nosed monkeys, Campbell's monkeys, gibbons, gorillas [20, 21, 55-58]	Coordination of group activities [21, 55-57]; Social and predatory contexts [20, 58]
Segmental concatenation	Combination of standalone meaningful segments and bound segments X = individual coding segment Y_N = behaviour N X Y_N = individual's identity + behaviour	*Ed* *go/look…* *Ed goes/looks…*	Diana monkeys, banded mongooses, dingoes [59-61]	Coordination of group activities [59-61]; Alerting function [62]
Affixation	Combination of meaningful call with a meaning-modifying bound affix X_N = meaning N Y = abstract meaning Y X_N Y = modified meaning N	*fold/pack…* *un-* *unfold/unpack…*	Campbell's monkeys [63, 64]	Alerting function [63, 64]
Semantically compositional structures	Combinations of meaningful calls into meaningful structure with derived meaning X = meaning X Y = meaning Y X Y = new meaning derived from X &Y	*caution!* *hot* *caution hot!*	Japanese great tits [66]	Alerting function [66]
Temporality	Number/rate of element repetitions/inter-element interval encodes categorical information a a = meaning X a a a = meaning Y a a = meaning Z	-	Colobus monkeys [78, 79] Mexican free-tailed bats [77]	Predatory context [78, 79] Behaviour specific [77]

Research aims

The objective of this dissertation is to further our understanding of the prevalence and diversity of vocal combinations outside of primates by investigating this ability in two highly social passerine birds: southern pied babblers (*Turdoides bicolor*) and chestnut-crowned babblers (*Pomatostomus ruficeps*). Given the extensive array of behaviours that require coordination, there has likely been a significant selective pressure on both species to evolve new and diverse call types. However, like most animal species [80], babblers are anatomically constrained in the number of different calls they can produce. Combining existing sounds and calls may therefore represent a potential mechanism applied by both species to increase the amount of information that can be encoded, facilitating the smooth management of a plethora of behaviours upon which the stability of these species' social and breeding system depend [81]. Such a comparative approach, in species distantly related to humans, will help elucidate the selective forces promoting the evolution of combinatorial communication, and ultimately insights into the evolution of our own language will be gained [2, 12].

Study species, sites and populations

Southern pied babbler and chestnut-crowned babbler

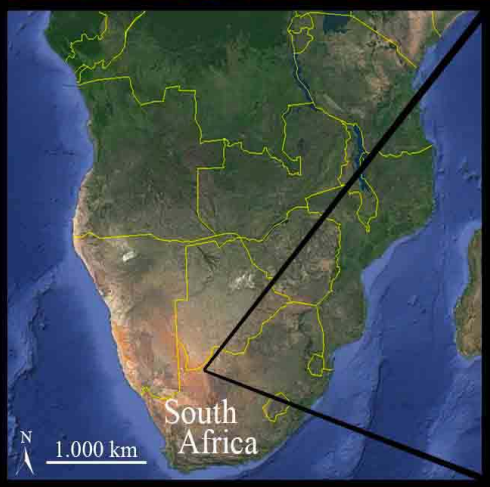
Southern pied babblers and chestnut-crowned babblers belong to two unrelated taxonomic groups designated as 'babblers': Old World babblers (Timaliidae) and Australo-Papuan babblers (Pomatostomidae) [82]. Both species are medium-sized, monomorphic, cooperative breeding passerines [83, 84] (Fig. 1 & 2). They spend a large amount of their time foraging on the ground feeding on invertebrates and small vertebrates, and are weak flyers with flights generally being short (15-40m) and low to the ground (0.5-3.5m) [85, 86]. Neither species produces songs, but instead a variety of alarm and social calls, which most likely evolved as a way to coordinate multiple behaviours inherent to the species' social system [25, 83, 87-92].

Southern pied babblers inhabit the semiarid grass- and open woodlands of the Kalahari Desert of southern Africa [88, 93]. The live in stable groups of up to 15 individuals consisting of a dominant breeding pair and subordinate individuals, which help to rear the offspring [83, 94, 95]. Individuals of a group forage and move in a highly cohesive manner [83]. Groups actively announce and defend their territory vocally and visually, but seldom physically, against neighbouring groups [24, 96].

Chestnut-crowned babblers are endemic to the arid/semiarid shrub- and open woodlands of south-eastern Australia [84]. During the non-breeding season they

Figure 1. From top left to bottom right: i) Group of southern pied babblers with the dominant pair vocalising. ii) Group of chestnut-crowned babblers scanning the environment (image curtsey of Jodie M. S. Crane). iii) Two foraging southern pied babblers digging in the substrate for prey. iv) A chestnut-crowned babbler foraging (image curtsey of Jodie M. S. Crane). v) Location[1] of and vi) habitat structure at the Pied Babbler Research Project. vi) Location[2] of and vii) habitat structure at the Chestnut-crowned Babbler Research Project.

Satellite image sources:
[1] Google earth V 7.1.5.1557 (May 20, 2015). Southern Africa. 20°43'27.73"S, 24°02'28.08"E, Eye alt 5438.87km. Digital Globe 2012. http://www.earth.google.com [August 11, 2016].
[2] Google earth V 7.1.5.1557 (May 20, 2015). Australia. 29°37'34.80"S, 141°24'27.73"E, Eye alt5431.93km. Digital Globe 2012. http://www.earth.google.com [August 11, 2016].


South
Africa
1.000 km

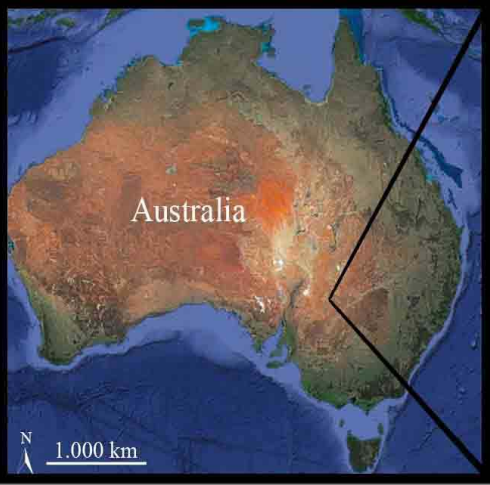

Australia
1.000 km

Figure 2. Southern pied babbler fledgling being fed by an adult bird, with another adult bird next to it preening.

live in groups of up to 23 individuals, which split up into smaller social groups during the breeding season, consisting of one breeding pair and subordinate helpers [84]. As groups fragment into subgroups they often overlap spatially during foraging [86]. Aggressive interactions between groups of chestnut-crowned babblers are rare [97].

Pied Babbler Research Project

A population of free-living, habituated southern pied babblers was studied at the Pied Babbler Research Project at Kuruman River Reserve in South Africa, established by Amanda R. Ridley in 2003 (26°58'S, 21°49'E) [88, 93] (Fig. 1). The study site covers an area of 33km^2, including plains, dunes, scrub- and open woodland, with a dry riverbed cutting through the reserve. The vegetation consists of perennial and annual grasses (*Aristida*, *Eragrostis*, *Schmidtia*, *Stipagrostis*), shrubs (*Grewia*, *Rhigozum*), and trees 3-8m in height (*Acicia*, *Boscia*) [93, 98]. Throughout the study period, the study population varied between 16-18 habituated, and 2-4 semi-habituated groups (i.e. groups that were in the progress of becoming fully habituated). Habituated individuals could be followed up to 1m, enabling close observations, audio recordings and experimental procedures. All individuals of the study population could be individually recognised through a unique combination of colour rings and a numbered metal ring [90]. The sex of individuals was assessed using DNA tests, and could further be determined through sex specific breeding behaviour and vocalisations [83, 92].

Chestnut-crowned Babbler Research Project

Wild, unhabituated chestnut-crowned babblers were studied at the Fowlers Gap Arid Zone Research Station in New South Wales, Australia (31°06'S, 141°42'E) [86] (Fig. 1). The study site covers an area of 64km^2 and is characterised by hills, creeks, and drainages, being sparsely vegetated with shrubs (*Maireana*, *Rhagodia*) and rows of trees (*Acacia*, *Casuarina*, *Heterodendrum*) [86]. The colour-ringed babbler population has been monitored since 2004 by Andrew F. Russell [97, 99]. For experimental procedures, birds were caught and housed in aviaries for a maximum time of 48h, and released again to their natal group after testing.

Thesis outline

The aim of this dissertation is to investigate to prevalence and variation of vocal combinations in the communication system of southern pied and chestnut-crowned babblers, ultimately identifying potential factors promoting the emergence of combinatorial mechanisms. Whilst chapter 1 provided an overview of combinatorial structures described in non-human vocal systems, the following chapters will investigate whether analogue forms also occur in the two babbler species, and specifically whether babblers produce basic phoneme-like or compositional constructs.

Chapter 2 investigates whether chestnut-crowned babblers reuse and combine acoustic elements to create context specific vocalisations. Two acoustically highly similar vocalisations are examined: the *AB*/'flight' call - a double-note call elicited during short flights, and the *BAB*/'prompt' call - a triple-note call elicited during nestling provisioning. Using natural observations, acoustic analyses and playback experiments the study investigates whether the two functionally distinct vocalisations are composed of the same acoustic elements (*A* & *B*), and whether the meaning differentiation between both vocalisations is the result of a single modification (presence/absence of the prompt call's first element), akin to a phonemic-like contrast in human language.

In chapter 3 the natural occurrence of call combinations in the vocal repertoire of the southern pied babbler is explored. Behavioural observations suggest that babblers produce 'alert' calls in response to low urgency threats, and 'recruitment' calls when recruiting group members during group travel. When encountering terrestrial predators, both vocalisations appear to be assembled into a 'mobbing sequence'. By applying a combination of acoustic analyses, predator presentations and playback experiments this work tests whether the sequence functions to recruit group members in dangerous situations, potentially linking information on the danger and the requested action, and whether the receiver parses the sequence in a compositional way.

While recruitment calls can be combined with other vocalisations to generate higher-order meaningful structures, the recruitment component itself seems to be characterised by additional within structure variation. Specifically, male southern pied babblers appear to produce two 'recruitment cry' variants, being similar in overall structure but differing in their internal structure, either comprised of repetitions of *A'*/single-note or *A'B'*/double-note motifs. Using acoustic analyses chapter 4 explores whether the same note type (i.e. *A* notes) is used across the two cry types. Behavioural observations and playback experiment will determine the cries' context specificity, postulating that both cries might function to recruit group members during group travel, with the internal structure modifying the form of recruitment from requesting approach to requesting following.

Chapter 5 addresses how temporal modifications within a sequence of repeated elements are applied by the southern pied babbler to generate distinct vocalisations. Previous work has shown that babblers produce 'purr' vocalisations when drawing young offspring to a food source [83, 88, 100]. While purrs are composed of, on average, around 14 element repetitions, babblers appear to produce a truncated version, the 'cluck' vocalisation, comprised of a limited number of 2-3 element repetitions. Acoustic analyses will investigate whether both vocalisations are composed of the identical sound type with only the number of repetitions differing between the two calls. Behavioural observations will serve to test whether clucks function to vocally coordinate imminent group movements between foraging sites.

Finally, chapter 6 will summarise the findings of this dissertation and propose potential factors that may have promoted the emergence of the combinatorial mechanisms, with a focus on the babbler systems. In conclusion, I will deduce the implications of this dissertation for our knowledge on the evolution of human language's generative system.

References

[1] Hockett CF. 1960 The Origin of Speech. *Sci. Am.* **203**, 88-111.
[2] Hauser MD, Chomsky N, Fitch WT. 2002 The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* **298**, 1569-1579. (doi:10.1126/science.298.5598.1569)
[3] Townsend SW, Manser MB. 2013 Functionally Referential Communication in Mammals: The Past, Present and the Future. *Ethology* **119**, 1-11. (doi:10.1111/eth.12015)
[4] Suzuki TN. 2016 Semantic communication in birds: evidence from field research over the past two decades. *Ecol. Res.* **31**, 307-319. (doi:10.1007/s11284-016-1339-x)
[5] Tomasello M. 2008 *Origin of human communication*. Cambridge, MA: MIT Press.
[6] Zuberbühler K. 2008 Audience effects. *Curr. Biol.* **18**, R189-R190. (doi:10.1016/j.cub.2007.12.041)
[7] Marler P, Dufty A, Pickert R. 1986 Vocal communication in the domestic chicken: II. Is a sender sensitive to the presence and nature of a receiver? *Anim. Behav.* **34**, 194-198. (doi:10.1016/0003-3472(86)90023-0)
[8] Hurford J. 2011 Linguistics from an Evolutionary Point of View. In *Handbook of the Philosophy of Science: Linguistics* (eds. R Kempson, M Fernando & N Ashe), pp. 473-498. Amsterdam: Elsevier Science.
[9] Bolhuis JJ, I. T, Chomsky N, Berwick RC. 2014 How Could Language Have Evolved? *PLoS Biol.* **12**, e1001934. (doi:10.1371/journal.pbio.100193)
[10] Humboldt Wv. 1999 *On Language. Translated by Peter Heath*. Cambridge: Cambridge University Press.
[11] Christiansen MH, Kirby S. 2003 Language evolution: consensus and controversies. *Trends Cogn. Sci.* **7**, 300-307. (doi:10.1016/s1364-6613(03)00136-0)
[12] Fitch WT. 2010 *The evolution of language*. Cambridge: Cambridge University Press.
[13] de Boer B, Sandler W, Kirby S. 2012 New perspectives on duality of patterning: Introduction to the special issue. *Lang. Cog.* **4**, 251-259. (doi:10.1515/langcog-2012-0014)
[14] Chomsky N. 1981 Knowledge of Language: Its Elements and Origins. *Philos. T. R. Soc. B* **295**, 223-234. (doi:10.1098/rstb.1981.0135)
[15] Hurford J. 2011 *The origins of grammar*. Oxford: Oxford University Press.
[16] Nowak MA, Krakauer DC. 1999 The evolution of language. *Proc. Natl. Acad. Sci. USA* **96**, 8028-8033. (doi:10.1073/pnas.96.14.8028)
[17] Nowak MA, Krakauer DC, Dress A. 1999 An error limit for the evolution of language. *Proc. R. Soc. B* **266**, 2131-2136. (doi:10.1098/rspb.1999.0898)
[18] Nowak MA, Plotkin JB, Jansen VAA. 2000 The evolution of syntactic communication. *Nature* **404**, 495-498. (doi:10.1038/35006635)
[19] Nowak MA, Komarova NL. 2001 Towards an evolutionary theory of language. *Trends Cogn. Sci.* **5**, 288-295. (doi:10.1016/S1364-6613(00)01683-1)
[20] Ouattara K, Lemasson A, Zuberbühler K. 2009 Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proc. Natl. Acad. Sci. USA* **106**, 22026-22031. (doi:10.1073/pnas.0908118106)
[21] Arnold K, Zuberbühler K. 2012 Call combinations in monkeys: Compositional or idiomatic expressions? *Brain Lang.* **120**, 303-309. (doi:10.1016/j.bandl.2011.10.001)
[22] Berwick RC, Okanoya K, Beckers GlJL, Bolhuis JJ. 2011 Songs to syntax: the linguistics of birdsong. *Trends Cogn. Sci.* **15**, 113-121. (doi:10.1016/j.tics.2011.01.002)
[23] Rendall D. 2013 Q&A: Cognitive ethology - inside the minds of other species. *BMC Biol.* **11**, 1-6. (doi:10.1186/1741-7007-11-108)
[24] Golabek KA. 2010 Vocal communication and the facilitation of social behaviour in the southern pied babbler (*Turdoides bicolor*), University of Bristol, PhD.
[25] Crane JMS, Savage JL, Russell AF. 2016 Diversity and function of vocalisations in the cooperatively breeding Chestnut-crowned Babbler. *Emu* **116**, 241. (doi:10.1071/mu15048)
[26] Bowling DL, Fitch WT. 2015 Do Animal Communication Systems Have Phonemes? *Trends Cogn. Sci.* **19**, 555-557. (doi:10.1016/j.tics.2015.08.011)
[27] Payne RS, McVay S. 1971 Songs of Humpback Whales. *Science* **173**, 585-597. (doi:10.1126/science.173.3997.585)
[28] Mitani JC, Marler P. 1989 A Phonological Analysis of Male Gibbon Singing Behavior. *Behaviour* **109**, 20-45.
[29] Catchpole CK, Slater TLB. 1995 *Bird Song: Themes and variations*. New York: Cambridge University Press.
[30] Geissmann T. 2002 Duet-splitting and the evolution of gibbon songs. *Biol. Rev.* **77**, 57-76. (doi:10.1017\S1464793101005826)
[31] Suzuki R, Buck JR, Tyack PL. 2005 Information entropy of humpback whale songs. *J. Acoust. Soc. Am.* **119**, 1849-1866. (doi:10.1121/1.2161827)
[32] Marler P, Slabbekoorn H. 1999 *Nature's Music: The science of birdsong*. New York: Academic Press.

[33] Holland J, Dabelsteen T, Paris AL. 2000 Coding in the song of the wren: importance of rhythmicity, syntax and element structure. *Anim. Behav.* **60**, 463-470. (doi:10.1006/anbe.2000.1529)
[34] Nousek AE, Slater PJ, Wang C, Miller PJ. 2006 The influence of social affiliation on individual vocal signatures of northern resident killer whales (*Orcinus orca*). *Biol. Lett.* **2**, 481-484. (doi:10.1098/rsbl.2006.0517)
[35] Payne K, Payne R. 2010 Large Scale Changes over 19 Years in Songs of Humpback Whales in Bermuda. *Z. Tierpsychol.* **68**, 89-114. (doi:10.1111/j.1439-0310.1985.tb00118.x)
[36] Antunes R, Schulz T, Gero S, Whitehead H, Gordon J, Rendell L. 2011 Individually distinctive acoustic features in sperm whale codas. *Anim. Behav.* **81**, 723-730. (doi:10.1016/j.anbehav.2010.12.019)
[37] Schulz TM, Whitehead H, Gero S, Rendell L. 2011 Individual vocal production in a sperm whale (*Physeter macrocephalus*) social unit. *Mar. Mam. Sci.* **27**, 149-166. (doi:10.1111/j.1748-7692.2010.00399.x)
[38] Kershenbaum A, Ilany A, Blaustein L, Geffen E. 2012 Syntactic structure and geographical dialects in the songs of male rock hyraxes. *Proc. R. Soc. B* **279**, 1-8. (doi:10.1098/rspb.2012.0322)
[39] Briefer EF, Rybak F, Aubin T. 2013 Does true syntax or simple auditory object support the role of skylark song dialect? *Anim. Behav.* **86**, 1131-1137. (doi:10.1016/j.anbehav.2013.09.019)
[40] King SL, Janik VM. 2013 Bottlenose dolphins can use learned vocal labels to address each other. *Proc. Natl. Acad. Sci. USA* **110**, 13216-13221. (doi:10.1073/pnas.1304459110)
[41] Janik V, Slater P. 1997 Vocal Learning in Mammals. *Adv. Stud. Behav.* **26**, 59-99. (doi:10.1016/s0065-3454(08)60377-0)
[42] Marler P. 2000 Origins of Music and Speech: Insights from Animals. In *The Origins of Music* (eds. WL Nils, B Merker & S Brown), pp. 31-48. Cambridge, MA: The MIT Press.
[43] Collier K, Bickel B, van Schaik CP, Manser MB, Townsend SW. 2014 Language evolution: syntax before phonology? *Proc. R. Soc. B* **281**, 20140263. (doi:10.1098/rspb.2014.0263)
[44] Hailman JP, Ficken MS, Ficken RW. 1985 The Chick-a-Dee Calls of Parus-Atricapillus - a Recombinant System of Animal Communication Compared with Written-English. *Semiotica* **56**, 191-224. (doi:10.1515/Semi.1985.56.3-4.191)
[45] Hailman JP. 1989 The Organization of Major Vocalizations in the Paridae. *Wilson Bull.* **101**, 305-343.
[46] Ficken MS, Hailman ED, Hailman JP. 1994 The Chick-a-Dee Call System of the Mexican Chickadee. *Condor* **96**, 70-82. (doi:10.2307/1369065)
[47] Suzuki TN. 2013 Communication about predator type by a bird using discrete, graded and combinatorial variation in alarm calls. *Anim. Behav.* **87**, 59-65. (doi:10.1016/j.anbehav.2013.10.009)
[48] Hailman JP, Ficken MS. 1986 Combinatorial animal communication with computable syntax: Chick-a-dee calling qualifies as 'Language' by structural linguistics. *Anim. Behav.* **34**, 1899-1901. (doi:10.1016/s0003-3472(86)80279-2)
[49] Freeberg TM, Lucas JR. 2002 Receivers respond differently to chick-a-dee calls varying in note composition in Carolina chickadees, *Poecile carolinensis*. *Anim. Behav.* **63**, 837-845. (doi:10.1006/anbe.2001.1981)
[50] Clucas BA, Freeberg TM, Lucas JYR. 2004 Chick-a-dee call syntax, social context, and season affect vocal responses of Carolina chickadees (*Poecile carolinensis*). *Behav. Ecol. Sociobiol.* **57**, 187-196. (doi:10.1007/s00265-004-0847-9)
[51] Freeberg TM, Lucas JR. 2012 Information Theoretical Approaches to Chick-a-dee Calls of Carolina Chickadees (*Poecile carolinensis*). *J. Comp. Psychol.* **126**, 68-81. (doi:10.1037/A0024906)
[52] Robinson JG. 1984 Syntactic Structures in the Vocalizations of Wedge-Capped Capuchin Monkeys, *Cebus olivaceus*. *Behaviour* **90**, 47-79.
[53] Cheney DL, Seyfarth RM. 1990 *How mokneys see the world*. Chicago: The University of Chicago Press.
[54] Hurford J. 2007 *The origins of meaning*. Oxford: Oxford University Press.
[55] Arnold K, Zuberbühler K. 2006 Language evolution: Semantic combinations in primate calls. *Nature* **441**, 303. (doi:10.1038/441303a)
[56] Arnold K, Zuberbühler K. 2008 Meaningful call combinations in a non-human primate. *Curr. Biol.* **18**, R202-R203. (doi:10.1016/j.cub.2008.01.040)
[57] Hedwig D, Mundry R, Robbins MM, Boesch C. 2014 Contextual correlates of syntactic variation in mountain and western gorilla close-distance vocalizations: Indications for lexical or phonological syntax? *Anim. Cogn.* **18**, 423-435. (doi:10.1007/s10071-014-0812-6)
[58] Clarke E, Reichard UH, Zuberbühler K. 2006 The Syntax and Meaning of Wild Gibbon Songs. *PLoS ONE* **1**, e73. (doi:10.1371/journal.pone.0000073)
[59] Candiotti A, Zuberbühler K, Lemasson A. 2012 Context-related call combinations in female Diana monkeys. *Anim. Cogn.* **15**, 327-339. (doi:10.1007/s10071-011-0456-8)
[60] Jansen DA, Cant MA, Manser MB. 2013 Segmental concatenation of individual signatures and context cues in banded mongoose (*Mungos mungo*) close calls. *BMC Biol.* **10**, 97. (doi:10.1186/1741-7007-10-97)

[61] Coye C, Zuberbühler K, Lemasson A. 2016 Morphologically structured vocalizations in female Diana monkeys. *Anim. Behav.* **115**, 97-105. (doi:10.1016/j.anbehav.2016.03.010)
[62] Deaux EC, Allen AP, Clarke JA, Charrier I. 2016 Concatenation of 'alert' and 'identity' segments in dingoes' alarm calls. *Sci. Rep.* **6**, 30556. (doi:10.1038/srep30556)
[63] Ouattara K, Lemasson A, Zuberbühler K. 2009 Campbell's Monkeys Use Affixation to Alter Call Meaning. *PLoS ONE* **4**, e7808. (doi:10.1371/journal.pone.0007808)
[64] Coye C, Ouattara K, Zuberbühler K, Lemasson A. 2015 Suffixation influences receivers' behaviour in non-human primates. *Proc. R. Soc. B* **282**, 20150265. (doi:10.1098/rspb.2015.0265)
[65] Greenberg J. 1957 *Essays in linguistics*. Chicago: University of Chicago Press.
[66] Suzuki TN, Wheatcroft D, Griesser M. 2016 Experimental evidence for compositional syntax in bird calls. *Nat. Commun.* **7**, 10986. (doi:10.1038/ncomms10986)
[67] Manser MB. 2001 The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency. *Proc. R. Soc. B* **268**, 2315-2324. (doi:10.1098/rspb.2001.1773)
[68] Templeton CN, Greene E, Davis K. 2005 Allometry of Alarm Calls: Black-Capped Chickadees Encode Information About Predator Size. *Science* **308**, 1934-1937. (doi:10.1126/science.1108841)
[69] Blumstein DT. 2007 The Evolution, Function, and Meaning of Marmot Alarm Communication. *Adv. Stud. Behav.* **37**, 371-401. (doi:10.1016/s0065-3454(07)37008-3)
[70] Mahurin EJ, Freeberg TM. 2008 Chick-a-dee call variation in Carolina chickadees and recruiting flockmates to food. *Behav. Ecol.* **20**, 111-116. (doi:10.1093/beheco/arn121)
[71] DuBois AL, Nowicki S, Searcy WA. 2009 Swamp sparrows modulate vocal performance in an aggressive context. *Biol. Lett.* **5**, 163-165. (doi:10.1098/rsbl.2008.0626)
[72] Courter JR, Ritchison G. 2010 Alarm calls of tufted titmice convey information about predator size and threat. *Behav. Ecol.* **21**, 936-942. (doi:10.1093/Beheco/Arq086)
[73] Sieving KE, Hetrick SA, Avery ML. 2010 The versatility of graded acoustic measures in classification of predation threats by the tufted titmouse *Baeolophus bicolor*: exploring a mixed framework for threat communication. *Oikos* **119**, 264-276. (doi:10.1111/j.1600-0706.2009.17682.x)
[74] Lemasson A, Ouattara K, Bouchet H, Zuberbuhler K. 2010 Speed of call delivery is related to context and caller identity in Campbell's monkey males. *Naturwissenschaften* **97**, 1023-1027. (doi:10.1007/s00114-010-0715-6)
[75] Rek P, Osiejuk TS. 2012 Temporal patterns of broadcast calls in the corncrake encode information arbitrarily. *Behav. Ecol.* **24**, 547-552. (doi:10.1093/beheco/ars196)
[76] Wheatcroft D. 2015 Repetition rate of calls used in multiple contexts communicates presence of predators to nestlings and adult birds. *Anim. Behav.* **103**, 35-44. (doi:10.1016/j.anbehav.2015.02.009)
[77] Bohn KM, Schmidt-French B, Ma ST, Pollak GD. 2008 Syllable acoustics, temporal patterns, and call composition vary with behavioral context in Mexican free-tailed bats. *J. Acoust. Soc. Am.* **124**, 1838-1848. (doi:10.1121/1.2953314)
[78] Schel AM, Tranquilli S, Zuberbühler K. 2009 The alarm call system of two species of black-and-white colobus monkeys (*Colobus polykomos* and *Colobus guereza*). *J. Comp. Psychol.* **123**, 136-150. (doi:10.1037/a0014280)
[79] Schel AM, Candiotti A, Zuberbühler K. 2010 Predator-deterring alarm call sequences in Guereza colobus monkeys are meaningful to conspecifics. *Anim. Behav.* **80**, 799-808. (doi:10.1016/j.anbehav.2010.07.012)
[80] Fitch WT. 2000 The evolution of speech: a comparative review. *Trends Cogn. Sci.* **4**, 258-267. (doi:10.1016/s1364-6613(00)01494-7)
[81] Hauser MD. 1997 *The Evolution of Communication*. Cambridge: MIT Press.
[82] Sibley CG, Ahlquist JE. 1985 The phylogeny and classification of the Australo-Papuan passerine birds. *Emu* **85**, 1-14. (doi:10.1071/MU9850001)
[83] Radford AN, Ridley AR. 2006 Recruitment Calling: A Novel Form of Extended Parental Care in an Altricial Species. *Curr. Biol.* **16**, 1700-1704. (doi:10.1016/j.cub.2006.06.053)
[84] Browning LE, Patrick SC, Rollins LA, Griffith SC, Russell AF. 2012 Kin selection, not group augmentation, predicts helping in an obligate cooperatively breeding bird. *Proc. R. Soc. B* **279**, 3861-3869. (doi:10.1098/rspb.012.1080)
[85] Ridley AR, Raihani NJ. 2006 Facultative response to a kleptoparasite by the cooperatively breeding pied babbler. *Behav. Ecol.* **18**, 324-330. (doi:10.1093/beheco/arl092)
[86] Portelli DJ, Barclay H, Russell DJF, Griffith SC, Russell AF. 2009 Social organisation and foraging ecology of the cooperatively breeding Chestnut-crowned Babbler (*Pomatostomus ruficeps*). *Emu* **109**, 153-162. (doi:10.1071/MU08065)
[87] Radford AN, Ridley AR. 2007 Individuals in foraging groups may use vocal cues when assessing their need for anti-predator vigilance. *Biol. Lett.* **3**, 249-252. (doi:10.1098/rsbl.2007.0110)
[88] Raihani NJ, Ridley AR. 2007 Adult vocalizations during provisioning: offspring response and postfledging benefits in wild pied babblers. *Anim. Behav.* **74**, 1303-1309. (doi:10.1016/j.anbehav.2007.02.025)

[89] Hollén LI, Bell MBV, Radford AN. 2008 Cooperative Sentinel Calling? Foragers Gain Increased Biomass Intake. *Curr. Biol.* **18**, 576-579. (doi:10.1016/j.cub.2008.02.078)
[90] Radford AN, Ridley AR. 2008 Close calling regulates spacing between foraging competitors in the group-living pied babbler. *Anim. Behav.* **75**, 519-527. (doi:10.1016/j.anbehav.2007.05.016)
[91] Bell MBV, Radford AN, Rose R, Wade HM, Ridley AR. 2009 The value of constant surveillance in a risky environment. *Proc. R. Soc. B* **276**, 2997-3005. (doi:10.1098/rspb.2009.0276)
[92] Golabek KA, Radford AN. 2013 Chorus-call classification in the southern pied babbler: multiple call types given in overlapping contexts. *Behaviour* **150**, 691-712. (doi:10.1163/1568539X-00003074)
[93] Hollén LI, Bell MBV, Russell A, Niven F, Ridley AR, Radford AN. 2011 Calling by Concluding Sentinels: Coordinating Cooperation or Revealing Risk? *PLoS ONE* **6**, e25010. (doi:0.1371/journal.pone.0025010)
[94] Ridley AR, Raihani NJ. 2007 Variable postfledging care in a cooperative bird: causes and consequences. *Behav. Ecol.* **18**, 994-1000. (doi:10.1093/beheco/arm074)
[95] Nelson-Flower MJ, Hockey PAR, O'Ryan C, Raihani NJ, du Plessis MA. 2011 Monogamous dominant pairs monopolize reproduction in the cooperatively breeding pied babbler. *Behav. Ecol.* **22**, 559-565. (doi:10.1093/beheco/arr018)
[96] Raihani NJ. 2008 Cooperation and conflict in pied babblers, University of Cambridge, PhD.
[97] Sorato E, Gullett P-R, Creasey MJS, Griffith SC, Russell AF. 2015 Plastic territoriality in group-living chestnut-crowned babblers: roles of resource value, holding potential and predation risk. *Anim. Behav.* **101**, 155-168. (doi:10.1016/j.anbehav.2014.12.012)
[98] Ridley AR, Thompson AM. 2011 Heterospecific egg destruction by Wattled Starlings and the impact on Pied Babbler reproductive success. *Ostrich* **82**, 201-205. (doi:10.2989/00306525.2011.618247)
[99] Russell AF. 2010 Breeding ecology of the Chestnut-crowned Babbler: a cooperative breeder in the desert. *Emu* **110**, 324-331. (doi:10.1071/MU10028)
[100] Raihani NJ, Ridley AR. 2008 Experimental evidence for teaching in wild pied babblers. *Anim. Behav.* **75**, 3-11. (doi:10.1016/j.anbehav.2007.07.024)

CHAPTER 2

Experimental evidence for phonemic contrasts in a nonhuman vocal system

The Public Library of Science Biology 13(6):e1002171

PLOS | BIOLOGY

RESEARCH ARTICLE

Experimental Evidence for Phonemic Contrasts in a Nonhuman Vocal System

Sabrina Engesser[1☯]*, Jodie M. S. Crane[2☯], James L. Savage[3], Andrew F. Russell[4,5‡]*, Simon W. Townsend[1‡]

1 Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland, **2** Department of Animal & Plant Sciences, University of Sheffield, Sheffield, United Kingdom, **3** Department of Zoology, University of Cambridge, Cambridge, United Kingdom, **4** Centre for Ecology & Conservation, College of Life & Environmental Sciences, University of Exeter, Penryn, Cornwall, United Kingdom, **5** Fowlers Gap Arid Zone Research Station, School of Biological, Earth & Environmental Sciences, University of New South Wales, Sydney, Australia

☯ These authors contributed equally to this work.
‡ AFR and SWT also contributed equally to this work.
* sabrina.engesser@ieu.uzh.ch (SE); a.russell@exeter.ac.uk (AFR)

Citation: Engesser S, Crane JMS, Savage JL, Russell AF, Townsend SW (2015) Experimental Evidence for Phonemic Contrasts in a Nonhuman Vocal System. PLoS Biol 13(6): e1002171. doi:10.1371/journal.pbio.1002171

Academic Editor: Asif A. Ghazanfar, Princeton University, UNITED STATES

Received: November 6, 2014

Accepted: May 8, 2015

Published: June 29, 2015

Data Availability Statement: All data files are available from Dryad http://dx.doi.org/10.5061/dryad.082v2.

Funding: Funding was provided by: Forschungskredit, University of Zurich (SE, #57191601, www.researchers.uzh.ch); Swiss National Science Foundation (SWT, #31003A_153065, http://www.snf.ch); Isle of Man Department of Education and University of Sheffield (JMSC); Natural Environment Research Council (studentship to University of Cambridge) (JLS), Standard Grant (AFR, #NE/K005766, www.nerc.ac.uk) (AFR); Australian Research Council Discovery

Abstract

The ability to generate new meaning by rearranging combinations of meaningless sounds is a fundamental component of language. Although animal vocalizations often comprise combinations of meaningless acoustic elements, evidence that rearranging such combinations generates functionally distinct meaning is lacking. Here, we provide evidence for this basic ability in calls of the chestnut-crowned babbler (*Pomatostomus ruficeps*), a highly cooperative bird of the Australian arid zone. Using acoustic analyses, natural observations, and a series of controlled playback experiments, we demonstrate that this species uses the same acoustic elements (*A* and *B*) in different arrangements (*AB* or *BAB*) to create two functionally distinct vocalizations. Specifically, the addition or omission of a contextually meaningless acoustic element at a single position generates a phoneme-like contrast that is sufficient to distinguish the meaning between the two calls. Our results indicate that the capacity to rearrange meaningless sounds in order to create new signals occurs outside of humans. We suggest that phonemic contrasts represent a rudimentary form of phoneme structure and a potential early step towards the generative phonemic system of human language.

Author Summary

A major question in language evolution is how its generative power emerged. This power, which allows the communication of limitless thoughts and ideas, is a result of the combinatorial nature of human language: meaningless phonemes can be combined to form meaningful words (*phonology*), and words can be combined to form higher-order, meaningful structures (*syntax*). While previous work has indicated the potential for animals to form syntax-like constructions, there exists little convincing evidence for a basic phonemic capacity in animals. Here, we demonstrate, using analyses combined with natural observations and playback experiments, that the cooperatively breeding chestnut-crowned babbler

Grant (AFR, DP1094295, www.arc.gov.au); and the Royal Society Fellowship scheme (AFR). The funders had no role in study design, data collection or analysis, decision to publish, or in the preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: DFA, Discriminant Function Analysis; GLMM, Generalized Linear Mixed Model.

reuses two meaningless acoustic elements to create two functionally distinct vocalizations. This result suggests the basic ability for phoneme structuring occurs outside of humans and provides insights into potential early evolutionary steps preceding the generative phonemic system of human language.

Introduction

The vast lexicons that characterise human languages are the product of physical and cognitive processes that guide the combination of a limited number of meaningless sounds (phonemes) in a variety of ways to generate new meaning [1,2]. In a simple example, the phonemes /k/, /æ/ and /t/ can be rearranged in different ways to create the words *cat* [kæt], *act* [ækt] or *tack* [tæk] [1]. Alternatively, the phoneme /k/ from the word *cat* can be eliminated to create the word *at* [æt], with the first position (i.e., presence or absence of the phoneme /k/) representing a phonemic contrast that generates the differentiation in meaning [3]. In all four arrangements, the meaningless phonemes maintain their acoustic identity across words, and this, paired with the arbitrary relationship between phoneme structure and word meaning, results in words with shared phonemes having distinct semantic content [4]. Such phoneme structure is a basic ingredient of word generation in human language, and when combined with the rules governing assemblages of meaningful words (a syntactic layer), provides much of language's generative power [5–7]. Despite the crucial role that phoneme structure plays in language, little is known about how such a capacity might have evolved [8–11]. Whilst comparative data from animal communication systems can elucidate early forms of language components, data demonstrating the critical rudiments of phoneme structures outside of humans is lacking.

Evidence that animals can employ a basic syntactical layer of language in their communication system has been provided in nonhuman primates. For example, Campbell's monkeys (*Cercopithicus cambelli*) produce two predator-specific alarm calls that are each modified in a predictable way into more general disturbance calls upon addition of the same suffix [12,13]. However, because the constituent calls are themselves meaningful (with the suffix carrying an abstract meaning in this case [14]), this, and equivalent findings [15,16], do not exemplify phoneme structure. Several candidates of phoneme-like structures in nonhuman animals have been proposed, but defining features are either lacking or have yet to be demonstrated [8,11,17]. One set of contenders comes from the songs of birds and mammals, in which meaningless elements are combined to create complex, higher-order structures [11,18,19]. However, experiments investigating behavioural responses to element reorganisation within songs are either lacking [18–21] or have not shown that such reorganisation confers a qualitative change in contextual meaning [22–24]. Another set includes calls produced in movement and alarm contexts. For example, parid birds can produce variable vocal sequences of apparently meaningless acoustic elements. However, in these cases, although call elements are commonly repeated or omitted, the required association between sequence structure and qualitative changes in informational content has not been demonstrated [25–29].

Using acoustic analysis, natural observations, and controlled playback experiments we provide evidence for rudimentary phoneme structure in the calls of chestnut-crowned babblers (*Pomatostomus ruficeps*) (see Materials and Methods), a 50 g, highly social, cooperatively breeding bird [30,31]. Observations over the past 10 years suggest that the repertoire of adult chestnut-crowned babblers consists of at least 15 discrete, context-specific vocalizations, of which three pairs appear to share sound elements, with the reused elements in each case being restricted to a specific pair of calls [32]. Here, we specifically focused on a single pair: a

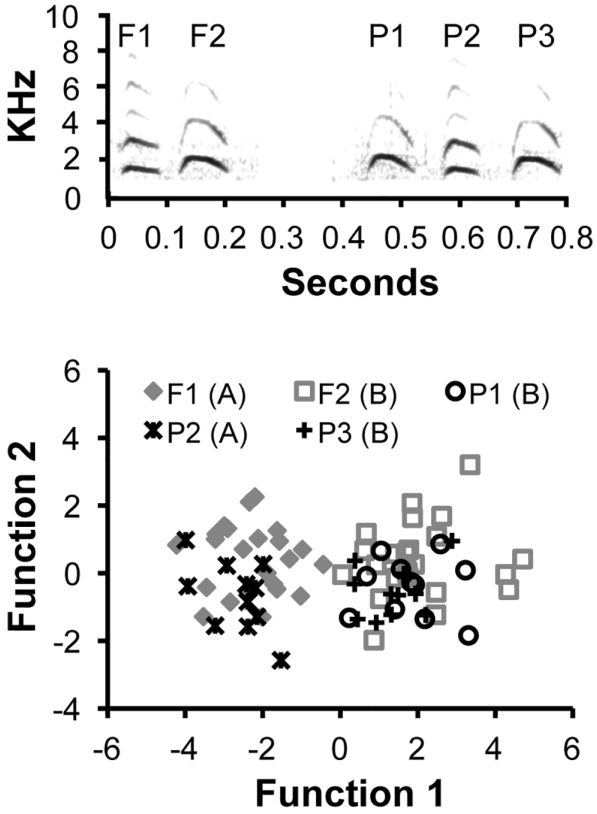


Fig 1. Flight and prompt call structure. (A) Spectrogram of double-element flight call (sequence F1 F2) and triple-element prompt call (sequence P1 P2 P3), taken from different individuals and groups. (B) Discriminant Function Analysis (DFA) output: function 1 explains 95% of the variance in element structure and primarily describes frequency range; function 2 explains the remaining 5% of variance and describes the contrast between start/end frequency (positive loadings) and frequency range (negative loadings) (Materials and Methods; S1 Text; S1 Table; S2 Table). F1 could not be reliably discriminated from P2 (34% errors: $T_{32} = 1.4$, $p = 0.2$) and nor could F2, P1, and P3 be discriminated from each other (27%–32% errors: F2 versus P1: $T_{32} = 0.7$, $p = 0.4$; F2 versus P3: $T_{32} = 1.4$, $p = 0.2$; P1 versus P3: $T_{20} = 0.2$, $p = 0.8$), but F1/P2 could be easily distinguished from F2/P1/P3 (all 0% errors) ($T_{32–44} = 14.1–22.9$; all p values < 0.001; S3 Table). Accordingly, flight calls and prompt calls follow *AB* and *BAB* construction, respectively.

doi:10.1371/journal.pbio.1002171.g001

double-element call produced during flight (flight calls, elements F1 and F2) and a triple-element call produced during nestling provisioning [33] (prompt calls, elements P1, P2, and P3) (Fig 1A). Importantly, the constituent elements within these calls appear to be contextually meaningless. For example, none of the elements is used as an individual call in isolation, suggesting that none can function to confer contextual information. Additionally, because none is used in combination with other call types, they cannot clearly operate to modify calls in a predictable way, as would be required of affixes [13]. First, we establish, using acoustic analyses, that the two calls comprise statistically equivalent acoustic elements. Second, we present natural observations showing that the two calls are context-specific, a prerequisite of reliable information transfer in animals. Finally, playbacks of natural, switched-element, and artificial calls in a standardised aviary environment confirm that the call elements are perceptibly equivalent and that element addition/elimination at one position creates a phoneme-like contrast, yielding the functional changes in meaning.

Results

Acoustic Analysis

Acoustic analyses were conducted to test whether prompt and flight calls are composed of statistically indistinguishable acoustic elements. To avoid problems of pseudo-replication arising from using calls of genetic relatives within groups [34], we analysed a single flight call and a single prompt call per group recorded (n = 23 flight, 11 prompt calls). Five parameters were extracted from the fundamental frequency of the resulting 79 elements: start and end frequency, frequency range, time to peak frequency, and element duration (S1 Text, S1 Table, and S2 Table). A Discriminant Function Analysis (DFA) demonstrated that the five elements across the two calls comprised just two independent acoustic structures (Fig 1B). Mahalanobis distances generated from the DFA revealed that F1 and P2 could not be reliably distinguished and neither could F2, P1, and P3 (all p values > 0.2), but that F1 and P2 could be distinguished easily from F2, P1, and P3 (all p values < 0.001) (Fig 1B and S3 Table). Thus, the two calls appear to comprise the same two distinct elements, with flight and prompt calls displaying *AB* and *BAB* patterns, respectively.

Natural Observations

Natural observations were conducted to quantify the context in which flight and prompt calls are produced. Natural flights were accompanied by flight calls in 274 of 450 observations (61%; n = 6 groups, 1 h/group), with all flights being short, low, and easily quantified. Similarly, hand-held releases following capture induced flight calls in 58 of 90 occasions (64%, n = 24 groups). No prompt calls were recorded in either set of observations, and flights/releases lack-ing flight calls were either silent or associated with alarm calls in response to observer presence. Finally, recordings from within nests in conjunction with automated nest entry-exit recorders revealed that 62% flights to/from nests were accompanied by flight calls (n = 140 visits, 7 groups) but rarely prompt calls (0.08% of nest visits), while 70% of nestling provisioning events were associated with prompt calls (n = 140 visits, 7 groups) and rarely flight calls (0.03% of nest visits). Additionally, in 97% of nest visits in which both flight and prompt calls were re-corded, individuals used only flight calls travelling to/from the nest and only prompt calls with-in nests (n = 60 visits, 7 groups). Thus, flight and prompt calls are highly context-specific, with the former maintaining group cohesion during movement [35] and the latter increasing the efficiency of food transfer to offspring by stimulating begging [33].

Playback Experiments

To verify experimentally that flight and prompt calls are context specific and are generated from rearrangement of the same acoustic elements, we performed playback experiments on 16 birds captured from 7 groups during periods of breeding. Each of the 16 birds received six playback trial-sets presented in a randomised order. Behavioural responses to two natural, two switched-element and two artificial calls were recorded in aviary compartments (2 x 2.5 x 2 m l x b x h) containing natural perches, foraging substrate, a view to the outside, and a recently used babbler nest (30x45 cm dome-shape, 6 cm diameter entrance) (Fig 2). The playback speaker was positioned out of view in a neighbouring compartment; birds had to look perpendicular to the speaker to look outside the aviary and in the opposite direction to look at the nest (S2 Text). Given our natural observations, we predicted flight calls would elicit increased observations to the outside and increased movement in anticipation of an incoming bird, while prompt calls would provoke greater nest attentiveness. Combined, these three behaviours comprised 61% of the activity budget in each trial (SD [standard deviation] = 23%; correlation

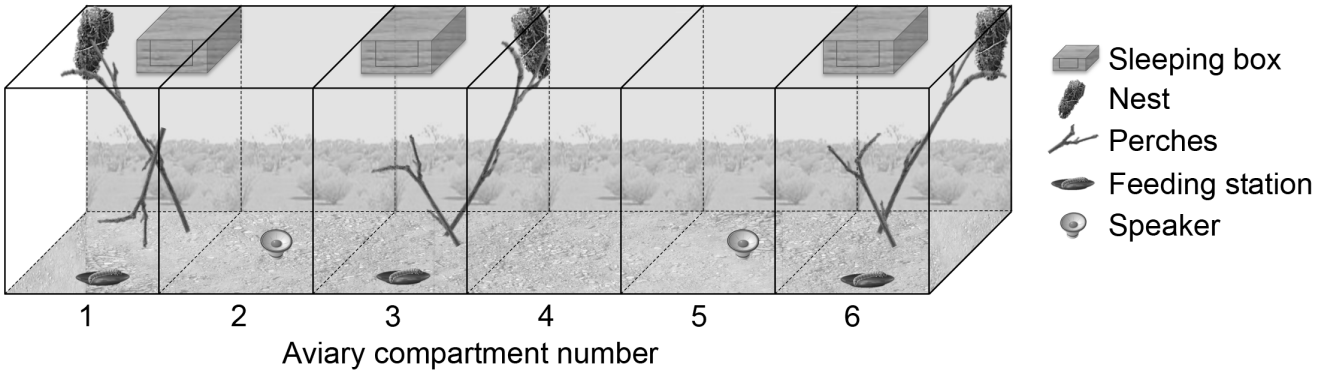


Fig 2. Schematic of aviary setup. The aviary consisted of six compartments: the back comprised metal meshing (1 cm^2), allowing the birds an outside view; the two sides were made of aluminium; and the front was specially designed perspex, allowing a one-way view from outside to inside. Occupied compartments contained natural perches, foraging substrate, a feeding station, babbler nest, and sleeping box, while unoccupied compartments contained the playback apparatus. Babbler nests are large (~45 x 30 cm), dome-shaped, with 6 cm diameter entrance hole, and robust. Babblers spent most of their time at mid-height; in all cases, relative to the speaker, birds had to look behind and up to look at the nest. Single birds used compartment 3 ($n = 2$), pairs of birds used compartments 1 and 3 ($n = 1$ pair) and trios used compartments 1, 3, and 6 ($n = 4$ trios) (S2 Text).

doi:10.1371/journal.pbio.1002171.g002

coefficients among these behaviours ranged from +0.1 to -0.3, indicating that time spent in one activity did not preclude time available for another).

Compared with natural prompt calls, natural flight call playbacks were associated with a 49% increase in the proportion of time spent looking outside (Generalized Linear Mixed Model [GLMM]: $\chi^2_1 = 11.8$, $p < 0.001$) and a 36% increase in time spent hopping/flying be-tween perches ($\chi^2_1 = 6.5$, $p = 0.02$). By contrast, during natural flight call playbacks, individuals spent 81% less time looking at the nest (2% of monitoring time) than during prompt call play-backs (15% of time) ($\chi^2_1 = 11.6$, $p < 0.001$) (Fig 3). Together, these results confirm the two calls are distinct and encode perceptible, context-specific information.

To test whether unmeasured acoustic variation dissociates the two calls [15], we played back switched-element versions of both calls to all 16 birds by generating flight calls from prompt elements P2 P3 and prompt calls from elements P1 F1 F2. The proportion of time birds spent engaged in the three behaviours of functional relevance were statistically equivalent between natural and switched-element flight calls (GLMM: all p values > 0.6; Fig 4A) as well as between natural and switched-element prompt calls (all p values > 0.3; Fig 4B). Additionally, there were no significant interactions between call type (flight versus prompt) and whether or not calls were natural or switched-element on behavioural responses (GLMM: all p values > 0.4). The absence of such interactions generated differences in behavioural responses to switched-element flight versus switched-element prompt calls of similar magnitude to those found in comparisons of natural calls (see Fig 4C versus Fig 3). Compared with switched-element prompt calls, switched-element flight calls were associated with 33% more time looking out, 33% more time in-movement, and 80% less time looking at the nest. Accordingly, it is improbable that any unmeasured acoustic differences between the elements of flight and prompt calls are responsible for the distinct responses, reinforcing our acoustic analyses that the calls comprise the same sound elements.

The results above suggest that the meaning-differentiating element between the two calls is P1. Before a phonemic-like system can be supported, two other interpretations require testing. First, element P1 might, by itself, be responsible for generating the contextual information carried by the prompt call, in which case, our results could be more akin to a syntactic, rather than phonemic, communicative system [12,13]. Second, the differences in response to flight calls

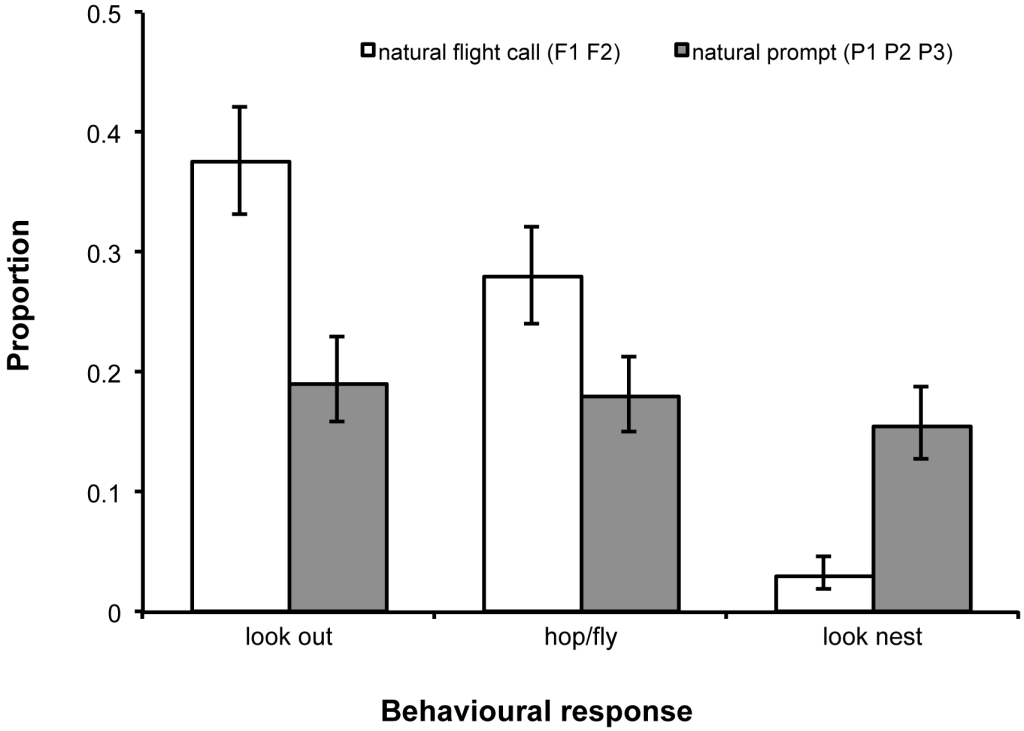


Fig 3. Responses to natural playbacks. Proportion of time spent engaged in three behaviours of functional relevance differed significantly during the playbacks of the two call types (see text). Figure shows back-transformed predicted means (± standard error [SE]) generated from three Generalized Linear Mixed Models (GLMM), in which the time engaged in each of the three activities (looking out of the aviary, in movement [hopping/flying], and looking at the nest) were fitted as three independent response terms. In each case, response terms were fitted to a binomial error structure with logit link function, time spent in camera view was fitted as the binomial denominator, call type (natural flight versus natural prompt) was fitted as a two-level factor, and individual identity nested within group identity were fitted as random terms.

doi:10.1371/journal.pbio.1002171.g003

versus prompt calls might arise from their differences in element number [36]. In this case, our results could represent stimulus intensity effects (triple-element prompt versus double-element flight call) or priming effects [12] (any acoustic element preceding a flight call results in a prompt-type response). To test these alternative interpretations, we presented two artificial stimuli to the 16 birds: element P1 alone and *CAB*, with the latter representing call elements P2 P3 (i.e., *AB*) preceded by an element (*C*) from chatter calls, a common call naturally repeated in mixed-element bouts and associated with excitement [32].

These two artificial stimuli elicited similar behavioural responses (all p values > 0.2; Fig 5A). First, they both generated relatively high look out and movement responses. One explana-tion lies with the fact that each is unnatural: impossible vocal scenarios have been shown to increase attentiveness behaviour in other contexts [37,38]. In support, separate analysis of the proportion of time spent looking around the aviary showed that general attentiveness behaviour during natural flight playbacks (mean ± SE = 16% ± 4%) was 36%, 47%, and 48% lower than during playbacks of *CAB*, P1, and natural prompts, respectively (GLMM: $\chi^2_3 = 10.6$, $p = 0.01$). Second, and more crucially, neither the P1- nor the CAB-stimulus elicited a hint of an elevated response in nest-attentiveness (Fig 5B). Like the flight call, P1 element and *CAB* playbacks were both associated with ca. 80% reductions in nest-attentiveness behaviour over natural prompt calls (Fig 5B). That neither the P1 element alone nor *CAB* elicits any increase in nest-attentiveness confirms that (a) P1 does not carry any nest-associated information in

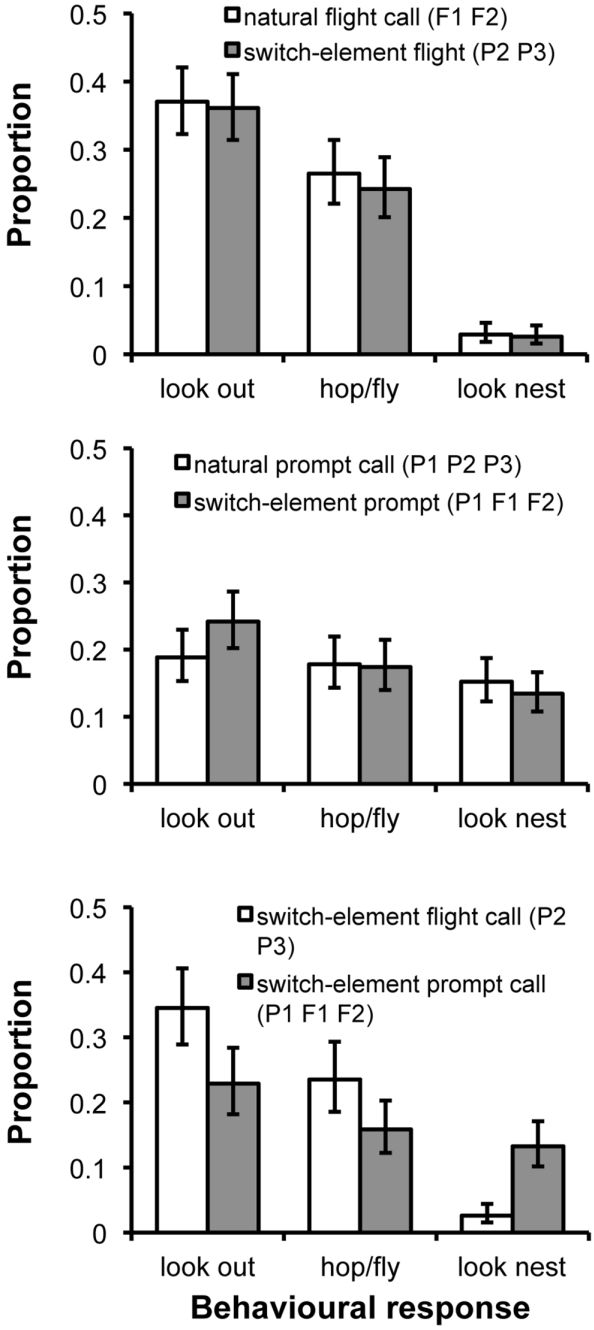


Fig 4. Responses to calls with and without reciprocal element exchange. Behavioural responses of functional relevance remained similar between (A) natural flight calls and switched-element flight calls comprising prompt call elements (natural versus switched-element comparisons: look out: $\chi^2_1 = 0.02$, $p = 0.9$; hop/fly: $\chi^2_1 = 0.2$, $p = 0.6$; look nest: $\chi^2_1 = 0.03$, $p = 0.9$) and (B) natural prompt calls and switched-element prompt calls using the two flight call elements (natural versus switched-element comparisons: look out: $\chi^2_1 = 0.2$, $p = 0.6$; hop/fly: $\chi^2_1 = 1.3$, $p = 0.3$; look nest: $\chi^2_1 = 0.01$, $p = 0.9$). (C) Behavioural responses to switched-element flight and switched-element prompt calls differed significantly or showed a non-significant tendency to do so (switched-element flight versus switched-element prompt call comparisons: look out: $\chi^2_1 = 5.7$, $p = 0.02$; hop/fly: $\chi^2_1 = 3.2$, $p = 0.09$; look nest: $\chi^2_1 = 10.0$, $p = 0.002$). Analyses were conducted as above (Fig 3) except that in (A) and (B) (which are shown separately for clarity) stimulus type (natural versus switched-

element) and its interaction with call type were added as additional fixed effects, while in (C), natural flight and prompt calls were replaced with switched-element ones. All figures show back-transformed predicted means (±SE).

doi:10.1371/journal.pbio.1002171.g004

isolation and (b) differential nest-attentiveness responses to flight and prompt calls are not derived from either stimulus intensity or priming effects. Thus, it is the presence or absence of

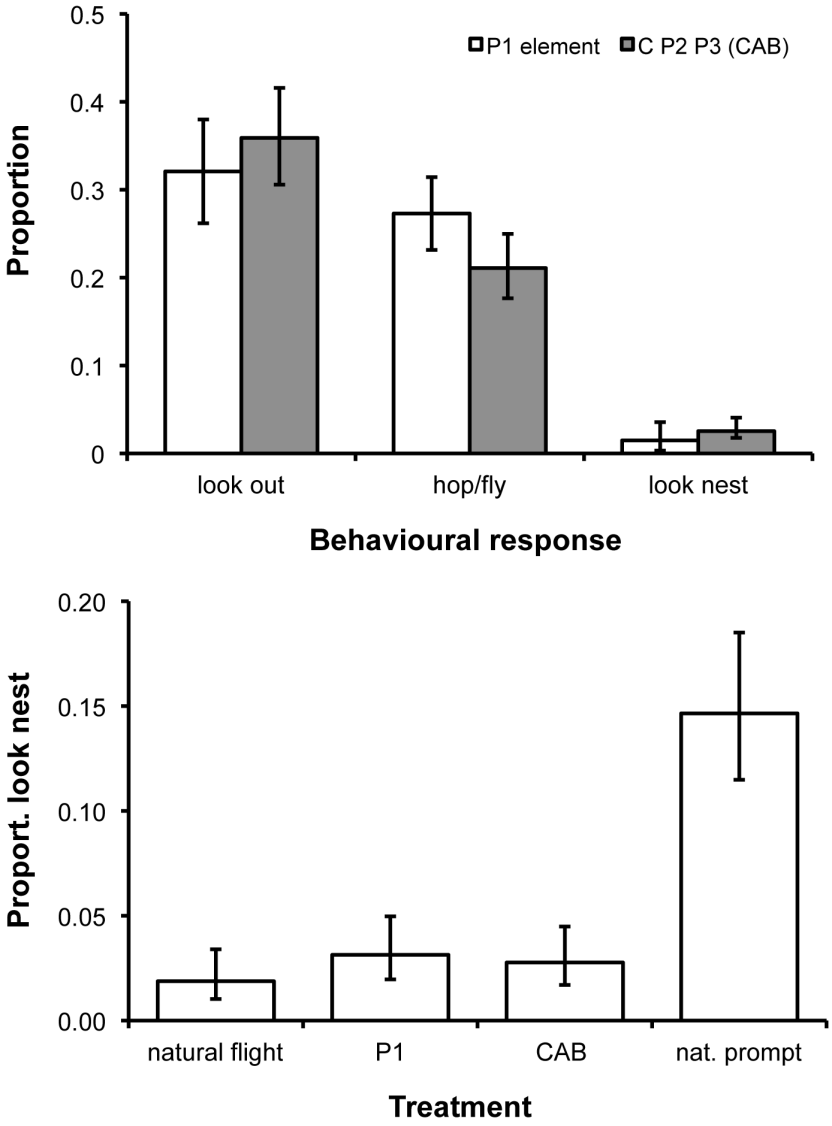


Fig 5. Behavioural responses to artificial calls. (A) During P1 element and *CAB* playbacks, individuals spent comparable proportions of time looking out ($\chi^2_1 = 0.2$, $p = 0.6$), in movement ($\chi^2_1 = 1.8$; $p = 0.2$), and looking at the nest ($\chi^2_1 = 0.5$; $p = 0.5$). (B) Neither P1 nor *CAB* playbacks provoked an increase in nest-attentiveness over natural flight calls, leading to proportions of time spent looking at the nest during these trials being substantially lower than those generated during natural prompt calls ($\chi^2_3 = 25.4$; $p < 0.001$). Figures show back-transformed predicted means (±SE) generated from GLMM analyses as outlined in Fig 3 (A) or in which the two-level factor was replaced with a four-level factor (B).

doi:10.1371/journal.pbio.1002171.g005

element P1 from the P2 P3 (or F1 F2) call sequence that appears integral to generating the qualitatively distinct meaning carried by the two calls.

Discussion

Phoneme structure represents a critical component of the vast lexicons in human languages, but a lack of suitably comparable evidence in animals has hindered our understanding of candidate selection pressures on, and early forms of, phoneme structure. Two related hypotheses have been proposed to explain the emergence of phonemic systems; both advocate a role of selection acting on increasing the capacity of vocal communication beyond that currently possible under an existing vocal repertoire. The here-named "enhanced-perception hypothesis" proposes that stringing together existing sounds in new ways reduces perception errors over the generation of new, but similar, sounds [39–41]. By contrast, the "vocal-constraints hypothesis" proposes that when the generation of new sounds is constrained [42], reusing pre-existing sounds in new combinatorial forms can provide an alternative solution to increasing communicative output [15,16]. Testing the predictions arising from these hypotheses represents a major challenge because human languages are generally too derived to address the pressures selecting for their emergence. Additionally, testing whether animals make perceptual mistakes for sounds that do not exist or are vocally constrained will be rarely feasible. A necessary first step in elucidating the pressures selecting for, and early forms of, phonemic structure is to address whether animals possess the capacity for generating functionally distinct vocalizations by rearranging contextually meaningless elements, and how such rearrangements are manifest.

Here, using acoustic analyses, natural observations, and playback experiments, we reveal that chestnut-crowned babblers use two acoustic elements (*A* and *B*) in different arrangements to create two functionally distinct vocalizations: flight calls (F1 F2, or *AB*) and prompt calls (P1 P2 P3, or *BAB*). The meaning differentiation between the two calls is not a result of the different number of elements or priming effects, but specifically the presence or absence of P1 (element *B*) at the head of the same call sequence. The fact that element P1 is both contextually meaningless on its own and meaning differentiating when used in combination with elements P2(F1) and P3(F2) signifies a phoneme-like contrast, with element *B* used in this position likely representing a phoneme-equivalent. To our knowledge, this is the first demonstration that animals have the basic capacity to use phoneme-like contrasts to derive qualitatively new meaning, a basic component of phoneme structuring. However, whether or not our results can also be interpreted as providing evidence for more advanced forms of phoneme structuring in an animal depends on two critical features.

First, in human languages, phoneme structure has potentially boundless generative power: the sum of derivable information is substantially greater than the number of its phonemic parts [1]. In contrast, the babbler vocal system that we describe is strictly bounded in its generative nature (i.e., two elements generate only two distinct calls). Part of the difference in human versus any nonhuman phonemic system will inevitably arise from vast differences in cognitive capacity [9]. Notwithstanding, cognitive capacity alone does not appear to be sufficient to explain differences in phonemic complexity and boundedness. For example, the sign language of the Al Sayyid Bedouin, an emerging language shared by deaf and hearing people of a small Israeli village, has been shown to have a fully functional and productive syntactic layer, but is so far characterized by only one phonological form [43,44]. Thus, when a phonemic layer emerges, even in human language, it appears initially to be finite and strongly bounded. This evidence suggests that the use of phonemic structure in communication should not be defined a priori by its complexity or boundedness, for it is likely that all phonemic systems evolve from simple beginnings like the one we describe here.

Second, the level of phonemic complexity used by babblers depends on the number of phoneme-equivalent entities in use. For example, whilst babblers generate a phonemic contrast by inserting the phoneme-like entity P1 before P2(F1) and P3(F2), whether or not P2 and F1 or P3 and F2 also represent phoneme-equivalent entities in the linguistic sense is equivocal. Unlike combinatoriality based on affixation rules or the generation of idioms, in which constituent parts have meaning [12,16], definitively testing whether all sound elements within call sequences of animals are contextually meaningless, and yet individually perceptible and meaning-differentiating, will be a major challenge. This is because any sound uttered by a conspecific can lead to a behavioural response irrespective of any perception of contextual meaning [38], and their limited vocal repertoires preclude investigation of whether distinct functional meaning is derived from the same meaningless elements in multiple different arrangements. A key component in discerning whether F1/P2 and F2/P3 are also phonemic depends on whether they represent a compound of two discrete elements, perceptible independently (i.e., *A* and *B*), or a holistic unit (i.e., *AB*). That the *B* element is phonemic in position P1 hints that *AB* is reducible, and hence F1/P2 and F2/P3 are probably also phoneme-like. However, this is an untested hypothesis at this stage, and we do not wish to speculate on whether chestnut-crowned babblers use more advanced forms of phoneme structure, beyond the identified use of a simple contrast, as part of their communication system.

Either way, we propose that the bounded use of phoneme-like contrasts in the vocal system of chestnut-crowned babblers represents a simple precursor of phoneme structuring that can elucidate how early forms of phonemic systems might emerge. For example, our results lead to the hypothesis that the addition or elimination of elements, i.e., basic phonemic contrasts (e.g., /kæt/ versus /æt/), might represent a simpler evolutionary step than complete element rearrangement (e.g. /kæt/ versus /tæk/), due to its reduced structural complexity. However, generating distinct contextual meaning through the former rather than latter process is likley to be more prone to perception errors, because it results in higher acoustic similarity. That babblers have opted for the more error-prone means of generating functionally distinct vocalizations, and done so by adding or eliminating a common element, is more supportive of a vocal-constraints hypothesis [15,16] than an enhanced-perception hypothesis [39–41]. Limiting the use of phonemic contrasts to short-range calls used in low-urgency, social contexts might be one way of reducing perception errors and mitigating associated costs when vocal constraints are operating.

In conclusion, the salient message here is that the basic capacity to generate qualitatively new meaning from rearranging contextually meaningless elements appears to exist outside of humans. One explanation is that for vocally constrained, highly social species, such as chestnut-crowned babblers, evolving new meaning by rearranging existing sounds offers a faster route to increasing communicative output than evolving new sounds. We hypothesise that reusing acoustic elements has facilitated the emergence of phoneme-like contrasts, which potentially drove sensitivity to phoneme structure or "phonemic awareness" in receivers [45,46]. The capacity to recognise vocalizations as sound constructs composed of smaller, meaningless elements, instead of a holistic unit, may have been the first step in the emergence of the elaborate phonemic systems seen in human languages. Further experiments are now required to determine exactly how babblers compute and perceive the elements from the two calls. More generally, further evidence for the use and manifestation of phonemic systems in animals is required; we propose that such systems will be most operant in the short-range communication of vocally constrained, social animals.

Materials and Methods

Study Site and Species

Ethics approval was provided by Macquarie University, Sydney, Australia (Number ARA 2013/025). The study was conducted on a population of wild, unhabituated chestnut-crowned babblers at the Fowlers Gap Arid Zone Research Station in far western New South Wales, Australia (141°42′E, 31°06′S). The population has been studied intensively since 2004. The habitat is characterised by low, open, chenopod shrubland, with trees largely confined to short, linear stands in drainage zones. Chestnut-crowned babblers (~50 g) are ground-foraging, weak-flying, and highly cooperative. During non-breeding they live in groups of 3–23 (mean ≈ 10) individuals, which then partially fragment into 1–4 units of 2–15 individuals (mean ≈ 6) for breeding. Non-breeders associate with those breeders to which they are most related and have substantial effects on their breeding success, primarily by reducing nestling starvation and facilitating additional reproductive attempts by the breeders. Further details on habitat and babbler socio-ecology are provided elsewhere [30,31,33,47–49]. All statistical analyses were performed in Genstat Release 17 (VSN International Ltd, Hemel Hempstead, UK, 2014). Data used in analyses and figure generation can be found in Dryad: http://dx.doi.org/10.5061/dryad.082v2 [50].

Context of Flight and Prompt Calls

We quantify the use of flight and prompt calls in three different contexts. First, in 2010, we used focal observations (1 h each on six groups from a distance of ~25 m) to determine the frequency with which the two calls are uttered during natural flights (n = 450 flights). Second, in 2011, we used a Fostex FR2-LE and wind-shielded Sennheiser ME67 shotgun microphone to record the vocalizations uttered during manual releases from cloth bags following capture (observers under a bedsheet; n = 90 releases from 25 groups). Third, in 2012, we fitted Yoga EM-400 mini tie-clip microphones to the wall of nests during nestling provisioning and recorded vocalizations using an Olympus LS-10 PCM or Fostex FR2-LE. To quantify the use of flight and prompt calls during flights to and from the nest, as well as during provisioning within the nest, we coupled the above nest-recording system with a transponder system, allowing the timing of bird entrances and exits to and from the nest to be determined [30,33,47]. Briefly, by inserting transponder tags (2 x 12 mm) into the flanks of the birds and fitting an antenna around the nest entrance linked to a TROVAN decoder, we were able to determine the use of the two calls within 5 s of entering and exiting the nest. Nest recordings were made 7 A.M.–4 P.M., in August–October, when broods were 1–12 days old. The first 20 nest visits within recording periods were used to quantify call-use at seven nests (time taken for 20 visits: 68–401 min.; n = 140 visits).

Acoustic Extractions and Statistical Analysis of Natural Calls

To quantify the resemblance among the five elements within and between double-element flight calls and triple-element prompt calls, we selected a single flight and prompt call recorded from each group during releases and nest recordings (sampling frequency of 44.1 kHz, 16 bits). Calls were selected randomly from those exhibiting no obscuring vocalizations, high signal-to-noise ratio and low background noise, and blindly with respect to the analyses. The elements of such calls (n = 23 double-element flight calls and 11 triple-element prompt calls) were then extracted using Raven Pro, version 1.4 (Bioacoustics Research Program, Cornell Lab of Ornithology, Ithaca, NY, 2011). Five parameters were extracted from the fundamental frequencies of the five elements in the two call types (start and end frequency, time to peak frequency,

frequency range, and element duration). All parameters were normalised when necessary and then centred to the mean and standardised by dividing the centralised mean values by 2-fold their standard deviation, allowing direct comparison of the importance of each parameters within and between models [51]. Two of the ten possible correlation coefficients among the five parameters were significantly positive: element length and frequency range ($r_p = 0.65$, $p < 0.001$); and start and end frequency ($r_p = 0.38$, $p = 0.002$) (S1 Table). Preliminary Analyses of Variance (ANOVA) (S1 Text) showed that time to peak frequency was statistically invariant across the five elements, and this was also the case for element durations after controlling for its correlation with frequency range. By contrast, start and end frequency, as well as frequency range, all varied between the elements, and for start and end frequency, this was the case after controlling for their correlation with each other. The three element parameters found to have significant independent effects on element structure (frequency range, start frequency, end frequency) were then used in a Discriminant Function Analysis (DFA) to determine acoustic similarity (S2 Table; S3 Table).

Playback Experiments: Test-Subject Selection and Housing

All playback experiments were conducted on wild birds captured in mist-nets on their territories during periods of breeding. Test subjects were chosen randomly from all adult birds captured (>6 months old), excluding the breeding female, without regard for sex and age. Depending on a test subject's group size, 1–3 individuals were removed (<30% of group members); resulting in 16 individuals from 7 different groups being tested. Removed birds were transported the 1–5 km immediately by car to aviaries on site at Fowlers Gap and released into separate aviary compartments (Fig 2). The aviaries consisted of six single compartments each of 2 m long, 2.5 m deep and 2 m high. Birds were housed singly, and each fed 20 mealworms every 2–3 h of daylight, delivered through a tube into each aviary compartment, of which 8–15 were typically consumed per bout. Birds gained a mean of 0.65 g (range = -3.1 to +4.8 g) during their time in the aviary; all birds were released near their original group less than 48 h after ini-tial capture, and were accepted back into their group without any signs of aggression [49].

Playback Experiments: Rationale, Call Recordings, and Playback Protocol

Our primary objective in this study was to test whether babblers used a phonemic contrast to generate qualitatively new information. For purposes of experimental rigour and analytical clarity, we chose a fully balanced design, with each bird being presented with the full set of selected playback stimuli. The drawback of presenting multiple stimuli to the same birds lies in the risk of habituation, leading to the generation of ambiguous results. For this reason, we decided to limit the number of playback trials to the absolute minimum number required to test for a phonemic contrast (i.e., six).

Our rationale for the six playback stimuli chosen was as follows. First, given the primary focus, the critical experiments needed to include natural and switched-element versions of both calls (i.e., amounting to four playback conditions). Second, because the acoustic analyses suggested that the only difference between the two calls derives from P1 in prompt calls, we deemed it key to test whether this element alone partially contributes to the overall meaning of the prompt call by eliciting an increased nest-attentiveness response compared with the flight call. If this were the case, we would have evidence of something more akin to a syntactic than phonemic system. Finally, because flight and prompt calls comprise two and three elements, respectively, we thought it essential to test for an influence of this difference in generating variation in nest attentiveness. We chose a stimulus including C1 P2 P3 because, again, we deemed

it most informative for the key aim to manipulate the one element that differs between the two calls (i.e., P1). The C1 element was taken from chatter calls: a common multi-element call uttered in bouts of several seconds in contexts of excitement or alarm [32]. The single C1 element was of comparable duration to the replaced P1 element.

The calls used in the playback experiments were obtained from natural recordings at the nest of six groups. In each case, a Sennheiser directional microphone (ME66/K6) connected to a Marantz solid-state recorder (PMD660, sampling frequency 48 KHz, 24 bits) was positioned within 1 m of a nest. Playbacks, including the construction of artificial calls (see below), were created with Adobe Audition CC (Version 6 Build 732, Adobe Systems), selecting high-quality calls (as above). Of the high-quality calls obtained, a single double-element flight call, triple-element prompt call, and a single element of the mixed-element chatter call were selected from each of the six groups (n = 18 calls). For each of the six groups from which recordings were obtained, the set of six playback stimuli were created, with each set including a natural flight call (F1 F2), a natural prompt call (P1 P2 P3), a switched-element flight call (P2 P3), a switched-element prompt call (P1 F1 F2), a P1 element stimulus (P1), and a triple-element stimulus (C1 P2 P3). In all cases, except for one, birds were tested with a new call-set played in randomized order, and birds never received a call-set from their own group. When elements for the generation of artificial and control calls were added and/or replaced, it was ensured that inter-element distance and amplitude matched the original call (Fig 1A). During each playback, a stimulus was repeated six times randomly distributed over 3.2–3.6 s; a break of at least 10 min was given for focal individuals to resume pre-stimuli behaviour before the initiation of another stimulus.

Playback experiments were conducted on the day following capture. Individuals of the same group were tested simultaneously with the same playback-set, but they were always housed separately and could not see each other (Fig 2 and S2 Text). Nevertheless, birds tested simultaneously could influence each other's behaviour if they reinforced (or countered) the playback experiment with their own vocalizations. This was not the case. In the 420 seconds of the playback experiment, not a single prompt call was uttered, and only 24 flight calls were given by the 14 individuals tested simultaneously, leading to a flight call rate of 0.28 per bird per 10 s trial. Additionally, of these 24, only ten were produced during natural or artificial flight call playbacks, all by two of the five groups. Finally, adding whether or not a flight call vocalization was uttered during the playbacks never impacted the explanatory power of the models (all p values > 0.8).

During testing, individuals were recorded using digital Sony handycams (HDR-CX220 and HDR-CX160) through a viewing hole to increase image clarity. Visual recordings of 10 s from playback onset were analysed frame by frame using Adobe Audition CC (Version 6 Build 732, Adobe Systems), with time (s) spent in camera view (mean = 9.4 s, range = 10–6 s), looking at the nest, looking outside (i.e., towards mesh wall), and in movement (hopping or flying) representing the primary parameters of interest, although general looking around behaviour was also recorded. Marker lists created in Adobe Audition were extracted into txt-files by using CueListTool (Version 1.7), and rates were calculated.

Playback Experiment: Statistical Analyses

Analyses of behavioural data arising from the playback experiments were conducted using Generalized Linear Mixed Models (GLMM), in which the time spent engaged in a given behaviour was fitted as the response term and the total amount of time spent in camera view was fitted as the binomial denominator. Explanatory terms included natural flight and natural prompt calls only (Fig 3); call type (flight or prompt), trial type (natural or switched-element), and their interaction (Fig 4A and 4B); switched-element flight and switched-element prompts

calls only (Fig 4C); and element P1 and *CAB* stimuli only (Fig 5A) or as a four-level factor with natural flight and natural prompt calls (Fig 5B). Additionally, the time spent in view was fitted as a covariate in a single movement analysis (Fig 3). In all GLMM analyses, individual identity nested within group identity were fitted as random terms. Doing so served two purposes: (1) it blocked the analyses by individual, effectively generating a more powerful repeated measures statistical design, and (2) it accounted for any lack of independence arising from testing birds from the same group simultaneously with the same playback stimuli. Regarding this potentially important latter issue: in all analyses, group identity was non-significant (all p values = 0.4–0.9), indicating that there was statistically equivalent variation in individual responses from the same group to the same playback stimuli as there was in individual responses from different groups to different playback stimuli.

Supporting Information

S1 Table. Correlations among call-element parameters.
(DOCX)

S2 Table. Discriminant Function Analysis (DFA) statistical output of call-element parameters.
(DOCX)

S3 Table. Mahalanobis D-squared distances generated from DFA.
(DOCX)

S1 Text. Statistical analysis of call-element parameters.
(DOCX)

S2 Text. Aviary set-up during playbacks.
(DOCX)

Acknowledgments

We thank Simon Griffith; Keith Leggett and the Dowling family for logistical support at Fowlers Gap; Elliot Capp, Matt Creasey, Marlijn Henskens, Andrea Liebl, Fumi Nomano, and Jeff Thomson for help with fieldwork; John Hunt for statistical advice; and Mark Aronoff, Balthasar Bickel, Elodie Briefer, Volker Dellwo, Marta Manser, Alex Thornton for helpful discussions and/or comments.

Author Contributions

Conceived and designed the experiments: SE JMSC JLS AFR SWT. Performed the experiments: SE JMSC JLS AFR. Analyzed the data: SE AFR. Wrote the paper: SE AFR SWT.

References

1. Hockett CF (1960) The Origin of Speech. Scientific American 203: 88–111. PMID: 13683472
2. Hauser MD (1997) The Evolution of Language. Cambridge: MIT Press.
3. de Boer B, Sandler W, Kirby S (2012) New perspectives on duality of patterning: Introduction to the special issue. Language and Cognition 4: 251–259. PMID: 24587833
4. Hurford J (2007) The origins of meaning. Oxford, New York: Oxford Univiversity Press.
5. Chomsky N (1981) Knowledge of Language: Its Elements and Origins. Philosophical Transactions of The Royal Society B: Biological Sciences 295: 223–234.
6. Nowak MA, Plotkin JB, Jansen VAA (2000) The evolution of syntactic communication. Nature 404: 495–498. PMID: 10761917

7. Hurford J (2011) Linguistics from an Evolutionary Point of View. In: Kempson R, Fernando M, Ashe N, editors. Handbook of the Philosophy of Science: Linguistics. Amsterdam: Elsevier Science.
8. Marler P (1977) The structure of animal communication sounds. Recognition of complex acoustic signals: report of Dahlem workshop. Berlin: Abakon Verlagsgesellschaft.
9. Hauser MD, Chomsky N, Fitch WT (2002) The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? Science 298: 1569–1579. PMID: 12446899
10. Yip MJ (2006) The search for phonology in other species. Trends in Cognitive Sciences 10: 442–446. PMID: 16931109
11. Berwick RC, Beckers GJL, Okanoya K, Bolhuis JJ (2012) A bird's eye view of human language evolution. Frontiers in Evolutionary Neuroscience 4: 5. doi: 10.3389/fnevo.2012.00005 PMID: 22518103
12. Ouattara K, Lemasson A, Zuberbühler K (2009) Campbell's Monkeys Use Affixation to Alter Call Meaning. PLoS ONE 4: e7808. doi: 10.1371/journal.pone.0007808 PMID: 19915663
13. Ouattara K, Lemasson A, Zuberbühler K (2009) Campbell's monkeys concatenate vocalizations into context-specific call sequences. Proceedings of the National Academy of Sciences 106: 22026–22031. doi: 10.1073/pnas.0908118106 PMID: 20007377
14. Collier K, Bickel B, van Schaik CP, Manser MB, Townsend SW (2014) Language evolution: syntax before phonology? Proceedings of the Royal Society B: Biological Sciences 281: 20140263. doi: 10.1098/rspb.2014.0263 PMID: 24943364
15. Arnold K, Zuberbühler K (2006) Language evolution: Semantic combinations in primate calls. Nature 441: 303. PMID: 16710411
16. Arnold K, Zuberbühler K (2008) Meaningful call combinations in a non-human primate. Current Biology 18: R202–R203. doi: 10.1016/j.cub.2008.01.040 PMID: 18334192
17. Catchpole CK, Slater TLB (1995) Bird Song: Themes and variations. New York: Cambridge University Press.
18. Payne RS, McVay S (1971) Songs of Humpback Whales. Science 173: 585–597. PMID: 17833100
19. Geissmann T (2002) Duet-splitting and the evolution of gibbon songs. Biological Reviews 77: 57–76. PMID: 11911374
20. Leger DW (2005) First Documentation of Combinatorial Song Syntax in a Suboscine Passerine Species. The Condor 107: 765–774.
21. Trillo PA, Vehrencamp SL (2005) Song types and their structural features are associated with specific contexts in the banded wren. Animal Behaviour 70: 921–935. PMID: 17173097
22. Holland J, Dabelsteen T, Paris AL (2000) Coding in the song of the wren: importance of rhythmicity, syntax and element structure. Animal Behaviour 60: 463–470. PMID: 11032649
23. Briefer EF, Rybak F, Aubin T (2013) Does true syntax or simple auditory object support the role of skylark song dialect? Animal Behaviour 86: 1131–1137.
24. Lachlan RF, Nowicki S (2015) Context-dependent categorical perception in a songbird. Proceedings of the National Academy of Sciences 112: 1892–1897. doi: 10.1073/pnas.1410844112 PMID: 25561538
25. Hailman JP (1989) The Organization of Major Vocalizations in the Paridae. The Wilson Bulletin 101: 305–343.
26. Ficken MS, Hailman ED, Hailman JP (1994) The Chick-a-Dee Call System of the Mexican Chickadee. The Condor 96: 70–82.
27. Freeberg TM, Lucas JR (2002) Receivers respond differently to chick-a-dee calls varying in note composition in Carolina chickadees, Poecile carolinensis. Animal Behaviour 63: 837–845.
28. Templeton CN, Greene E, Davis K (2005) Allometry of Alarm Calls: Black-Capped Chickadees Encode Information About Predator Size. Science 308: 1934–1937. PMID: 15976305
29. Suzuki TN (2013) Communication about predator type by a bird using discrete, graded and combinatorial variation in alarm calls. Animal Behaviour 87: 59–65.
30. Browning LE, Patrick SC, Rollins LA, Griffith SC, Russell AF (2012) Kin selection, not group augmentation, predicts helping in an obligate cooperatively breeding bird. Proceedings of the Royal Society B: Biological Sciences 279: 3861–3869. PMID: 22787025
31. Rollins LA, Browning LE, Holleley CE, Savage JL, Russell AF, et al. (2012) Building genetic networks using relatedness information: a novel approach for the estimation of dispersal and characterization of group structure in social animals. Molecular Ecology 21: 1727–1740. doi: 10.1111/j.1365-294X.2012.05492.x PMID: 22335253
32. Crane JMS (2014) Vocal communication in the cooperatively breeding chestnut-crowned babbler [PhD thesis]: University of Sheffield, PhD.

33. Young CM, Browning LE, Savage JL, Griffith SC, Russell AF (2013) No evidence for deception over allocation to brood care in a cooperative bird. Behavioral Ecology 24: 70–81.

34. Sharp SP, McGowan A, Wood MJ, Hatchwell BJ (2005) Learned kin recognition cues in a social bird. Nature 434: 1127–1130. PMID: 15858573

35. King BR (1980) Social-Organization and Behavior of the Grey-Crowned Babbler Pomatostomus-Temporalis. Emu 80: 59–76.

36. Slocombe KE, Kaller T, Call J, Zuberbuhler K (2010) Chimpanzees Extract Social Information from Agonistic Screams. PLoS ONE 5: e11473. doi: 10.1371/journal.pone.0011473 PMID: 20644722

37. Cheney DL, Seyfarth RM, Silk JB (1995) The responses of female baboons (Papio cynocephalus ursinus) to anomalous social interactions: evidence for causal reasoning? Journal of Comparative Psychology 109: 134–141. PMID: 7758289

38. Townsend SW, Allen C, Manser MB (2012) A simple test of vocal individual recognition in wild meerkats. Biology Letters 8: 179–182. doi: 10.1098/rsbl.2011.0844 PMID: 21992821

39. Nowak MA, Krakauer DC, Dress A (1999) An error limit for the evolution of language. Proceedings of the Royal Society B: Biological Sciences 266: 2131–2136. PMID: 10902547

40. Nowak MA, Krakauer DC (1999) The evolution of language. Proceedings of the National Academy of Sciences 96: 8028–8033. PMID: 10393942

41. Pinker S, Jackendoff R (2005) The faculty of language: what's special about it? Cognition 95: 201–236. PMID: 15694646

42. Janik VM, Slater PJ (2000) The different roles of social learning in vocal communication. Animal Behaviour 60: 1–11. PMID: 10924198

43. Sandler W, Aronoff M, Meir I, Padden C (2011) The gradual emergence of phonological form in a new language. Natural Language & Linguistic Theory 29: 503–543.

44. Sandler W, Aronoff M, Padden C, Meir I (2014) Language emergence: Al-Sayyid Bedouin Sign Language. In: Enfield NJ, Kockelman P, Sidnell J, editors. The Cambridge Handbook of Linguistic Anthropology. Cambridge: Cambridge University Press. pp. 246–278.

45. Anthony JL, Francis DJ (2005) Development of Phonological Awareness. Current Directions in Psychological Science 14: 255–259.

46. Pepperberg IM (2007) Grey parrots do not always 'parrot': the roles of imitation and phonological awareness in the creation of new labels from existing vocalizations. Language Sciences 29: 1–13.

47. Browning LE, Young CM, Savage JL, Russell DJF, Barclay H, et al. (2012) Carer provisioning rules in an obligate cooperative breeder: Prey type, size and delivery rate. Behavioral Ecology and Sociobiology 66: 1639–1649.

48. Sorato E, Gullett P-R, Creasey MJS, Griffith SC, Russell AF (2015) Plastic territoriality in group-living chestnut-crowned babblers: roles of resource value, holding potential and predation risk. Animal Behaviour 101: 155–168.

49. Nomano FY, Browning LE, Savage JL, Rollins LA, Griffith SC, et al. (2015) Unrelated helpers neither signal contributions nor suffer retribution in chestnut-crowed babblers. Behavioral Ecology. doi: 10.1093/beheco/arv023

50. Engesser S (2015). Data from: Experimental evidence for phonemic contrasts in a non-human vocal system. Dryad Digital Repository. Openly available via http://dx.doi.org/10.5061/dryad.082v2.

51. Schielzeth H (2010) Simple means to improve the interpretability of regression coefficients. Meth Ecol Evol 1: 103–113.

S1 Table. Pearson correlation coefficients (r_p) and significant values in parentheses for the five acoustic parameters measured from the two calls.

	EL	SF	EF	FR	TPF
Element length (EL)	-				
Start frequency (SF)	-0.10 (0.53)	-	-	-	
End frequency (EF)	-0.087 (0.55)	0.38 (**0.002**)	-	-	
Frequency range (FR)	0.65 (**<0.001**)	0.12 (0.37)	0.067 (0.95)	-	
Time peak frequency (TPF)	0.30 (0.69)	-0.24 (0.11)	0.30 (0.064)	-0.23 (0.15)	-

S2 Table. Discriminant function output, based on the inclusion of frequency range, start frequency and end frequency.

	DF1	DF2
Eigenvalue	4.66	0.24
% variance	95	5
Statistics (χ^2)	145.6, *DF*=12, *P*<0.001	17.3, *DF*=6, P=0.008
Standardised vector loadings of traits		
Start frequency	0.15	0.68
End frequency	0.16	0.82
Frequency range	0.78	-0.62

S3 Table. Mahalanobis D-squared distances of among element comparisons in DFA and associated probability values in parentheses based on t distributions with *DF*=*N*-2.

Element	F1	F2	P1	P2	P3
Flight 1 (F1)	-				
Flight 2 (F2)	18.4 (**<0.001**)	-	-	-	
Prompt 1 (P1)	17.4 (**<0.001**)	0.7 (0.4)	-	-	
Prompt 2 (P2)	1.4 (0.2)	22.9 (**<0.001**)	20.1 (**<0.001**)	-	
Prompt 3 (P3)	14.1 (**<0.001**)	1.4 (0.2)	0.2 (0.8)	16.1 (**<0.001**)	-

Supplementary Methods 1: Analysis of Call-Element Parameters

We first determined whether the five extracted acoustic parameters captured significant variation among the five call elements (F1, F2, P1, P2, P3) using a multivariate analysis of variance (MANOVA). To this end, we fitted the five call parameters as response terms, the 5-level factor pertaining to the five elements as the fixed effect, and the individual to which the elements belonged as a blocking function. Overall, there was a highly significant effect of element type on the response terms considered, indicating that our parameters captured meaningful variation in element structure ($F_{20,124}=9.57$, $P<0.001$; Wilks $\lambda=0.045$). Next, we identified those element parameters that independently captured variation in call elements using univariate ANOVAs, in which any correlated element parameters (S1 Table) were fitted as a co-variate. These analyses revealed that three call parameters tended to vary independently among the five elements: frequency range ($F_{4,40}=25.05$, $P<0.001$ controlling for element duration), start frequency ($F_{4,40}=2.58$, $P=0.058$ controlling for end frequency) and end frequency ($F_{4,40}=3.38$, $P=0.018$ controlling for start frequency). Element identity failed to account for significant variation in time to peak frequency ($F_{4,41}=0.12$, $P=0.98$) or element duration after controlling for frequency range ($F_{4,40}=1.44$, $P=0.24$).

Fitting the 3 significant (or near significant) acoustic parameters into a Discriminant Function Analysis (DFA) revealed that Discriminant Function 1 (DF1) explained 95% of the variance ($\chi 2=145.56$, $DF=12$, $P<0.001$) and Discriminant Function 2 (DF2) explained the remaining 5% ($\chi 2=17.27$, $DF=6$, $P=0.008$). Consideration of the eigenvalues and standardised vector loadings suggested that DF1 is heavily weighted to frequency range compared with the other two traits, while DF2 is positively weighted to start and end frequency and negatively to frequency range; thus DF2 explains the contrast between frequency range and start/end frequency (S2 Table). Fig. 1B shows the dominant effect of function 1 over function 2 on element discrimination (i.e. greater discrimination along the x than y axis), suggesting a primary effect of frequency range on element discrimination.

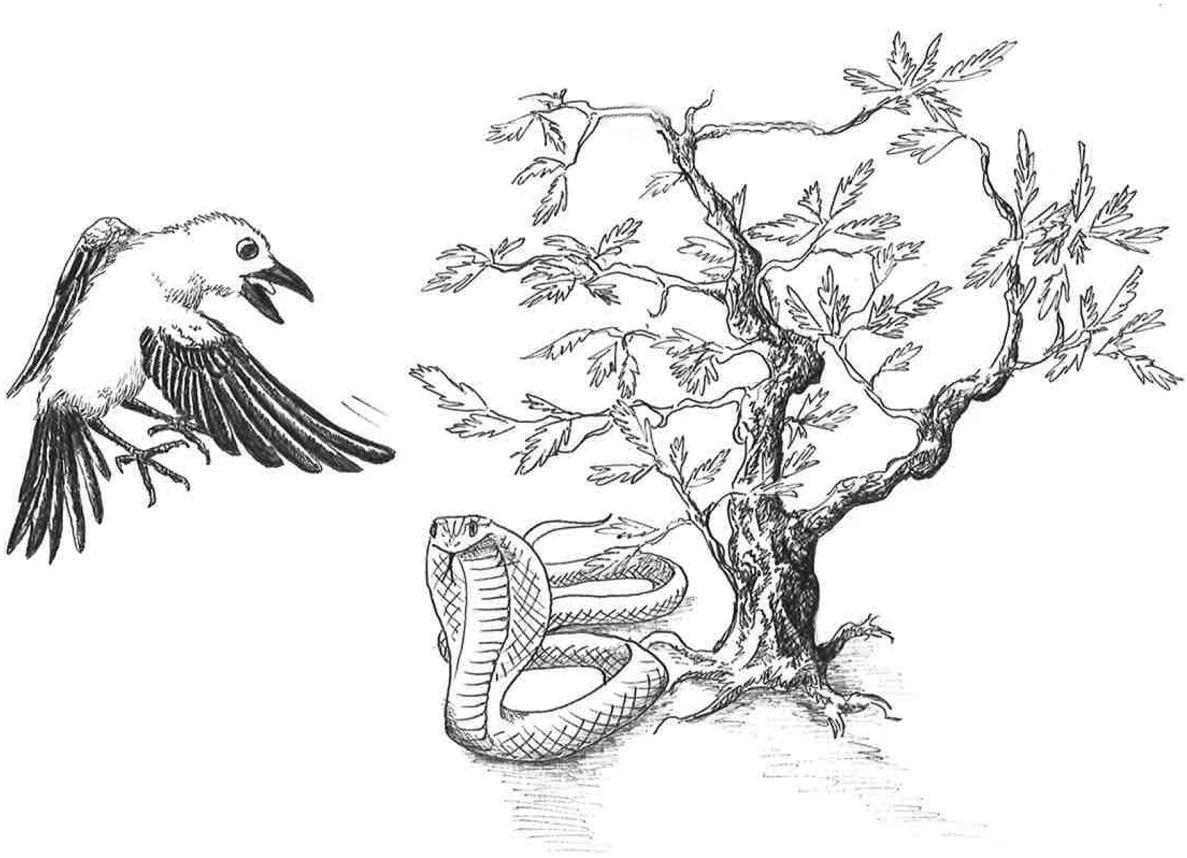
Supplementary Methods 2: Bird Locations During Aviary Playbacks

If one individual was tested, playbacks were broadcasted using a XMI X-mini II speaker placed in the empty aviary compartment 2 next to the compartment 3 of the focal bird. If two individuals were tested, the focal individuals were kept in two different compartments with the speaker in an empty compartment in-between (i.e. bird A in compartment 1; speaker in compartment 2 and bird B in compartment 3). When a third individual was tested, the two first birds were distributed in compartments as described for two birds; the third was kept in the compartment furthest away (i.e. compartment 6). In this case, playbacks were broadcasted with a Sony SRS-A27 stereo speaker system with one speaker again in the empty compartment in-between the first two test subjects (compartment 2), and the second in the empty compartment next to the third individual (compartment 5).

CHAPTER 3

Meaningful call combinations and compositional processing in the southern pied babbler

Proceedings of the National Academy of Sciences 113(21):5976-5981



Meaningful call combinations and compositional processing in the southern pied babbler

Sabrina Engesser[a,b,1], Amanda R. Ridley[a,c,d], and Simon W. Townsend[a,b,e]

[a]Pied Babbler Research Project, Kuruman River Reserve, Northern Cape 8467, South Africa; [b]Animal Behaviour, Department of Evolutionary Biology and Environmental Studies, University of Zurich, 8057 Zurich, Switzerland; [c]Centre for Evolutionary Biology, School of Animal Biology, The University of Western Australia, Crawley, WA 6009, Australia; [d]Percy FitzPatrick Institute, University of Cape Town, Rondebosch 7701, South Africa; and [e]Department of Psychology, University of Warwick, Coventry CV4 7AL, United Kingdom

Edited by Raghavendra Gadagkar, Indian Institute of Science, Bangalore, India, and approved April 5, 2016 (received for review January 26, 2016)

Language's expressive power is largely attributable to its compositionality: meaningful words are combined into larger/higher-order structures with derived meaning. Despite its importance, little is known regarding the evolutionary origins and emergence of this syntactic ability. Although previous research has shown a rudimentary capability to combine meaningful calls in primates, because of a scarcity of comparative data, it is unclear to what extent analog forms might also exist outside of primates. Here, we address this ambiguity and provide evidence for rudimentary compositionality in the discrete vocal system of a social passerine, the pied babbler (*Turdoides bicolor*). Natural observations and predator presentations revealed that babblers produce acoustically distinct alert calls in response to close, low-urgency threats and recruitment calls when recruiting group members during locomotion. On encountering terrestrial predators, both vocalizations are combined into a "mobbing sequence," potentially to recruit group members in a dangerous situation. To investigate whether babblers process the sequence in a compositional way, we conducted systematic experiments, playing back the individual calls in isolation as well as naturally occurring and artificial sequences. Babblers reacted most strongly to mobbing sequence playbacks, showing a greater attentiveness and a quicker approach to the loudspeaker, compared with individual calls or control sequences. We conclude that the sequence constitutes a compositional structure, communicating information on both the context and the requested action. Our work supports previous research suggesting combinatoriality as a viable mechanism to increase communicative output and indicates that the ability to combine and process meaningful vocal structures, a basic syntax, may be more widespread than previously thought.

call combination | compositionality | syntax | language evolution | southern pied babbler

Syntax is often considered one of the key defining features of human language (1). Through combining meaningful words together, larger sequences with related, compositional meaning can be constructed (2). One consequence of such productive compositional syntax in humans is that, with a finite inventory of words, an infinite range of ideas and concepts can be communicated (2, 3). Despite the central role that syntax plays in determining language's generativity, very little is known about its evolutionary origins or early, rudimentary forms (4, 5). Elucidating the proto forms of compositional syntax, although nontrivial (5, 6), represents a key step in understanding the evolution of language more holistically.

One means of investigating early forms and function of compositionality is to assess analog examples in animals (5, 7). Indeed, recent observational and experimental work on two related guenon monkeys has shown the propensity to combine context-specific, "meaningful" signals into sequences that resemble compositional structures in language (8–10). Male Campbell's monkeys (*Cercopithecus campbelli*), for example, produce predator-specific alarm calls that can be affixed with an acoustic modifier (8, 11). The affix acts to alter the "meaning" of the alarm calls in a predictable way, transforming them into general disturbance calls (8, 11, 12). Similarly, male putty-nosed monkeys (*Cercopithecus nictitans*) combine two predator-specific alarm calls into a higher-order sequence (9, 13). Although the two calls are generally associated with the presence of aerial or terrestrial predators, the resultant combination initiates group movement in nonpredatory contexts (9, 13). Given the discrepancies between the responses elicited by the individual calls and the sequence, it remains unclear whether the putty-nosed monkey call sequence represents a basic form of compositional syntax or rather a combinatorial syntax, where the meaning of the whole is not directly related to the parts, akin to idiomatic expressions in language (i.e., "kick the bucket" for dying) (10, 13, 14). The existence of such "semantic combinations" (13) in primates has nevertheless been argued to support an evolutionarily ancient origin of human syntax rooted within the primate lineage (8, 15). However, it is unclear whether similar call concatenations and compositional processing of information might also exist in other lineages (see ref. 14 for review) and if so, whether they take analogous forms and serve analogous functions (1).

The last 50 y of comparative research have shown that a number of nonprimate animals, particularly songbirds, are capable of stringing sounds together into larger, often more structurally complex sequences (16–18). However, there is no indication that any of these song sequences are compositional in structure, because the individual sounds composing the songs of birds and other animals do not convey any independent meaning (16–18), ultimately precluding any attempt to test for proto-syntactic abilities in these species in the first place. Although the

Significance

Human language is syntactic in its nature: meaningful words are assembled into larger meaningful phrases or sentences. How unique this ability is to humans remains surprisingly unclear. A considerable body of work has indicated that birds are capable of combining sounds into large, elaborate songs, but there is currently no evidence suggesting that these structures are syntactic. Here, we provide important evidence for this ability in a highly social bird. Specifically, pied babblers combine two functionally distinct vocalizations into a larger sequence, the function of which is related to the function of its parts. Our work adds important evidence to the variation and distribution of combinatorial vocal mechanisms outside humans and provides insights into potentially early forms of human syntactic communication.

Author contributions: S.E., A.R.R., and S.W.T. designed research; S.E. performed research; S.E. analyzed data; and S.E., A.R.R., and S.W.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Data has been deposited in Dryad (dx.doi.org/10.5061/dryad.ds0kb).

[1]To whom correspondence should be addressed. Email: sabrina.engesser@ieu.uzh.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1600970113/-/DCSupplemental.

absence of compositional structure in songs might suggest that syntactic abilities are potentially confined to the primate lineage (8, 15), it may also be an artifact of limited focus on bird vocal systems other than song that are more likely to support the capacity for syntax (19).

Here, we address this ambiguity through investigating the prevalence of compositional vocal sequences in a highly social, nonsinging passerine bird that possesses a discrete vocal system: the cooperatively breeding southern pied babbler (*Turdoides bicolor*) (20, 21). Pied babblers are territorial and live in stable groups of 3–15 individuals (22). Reproduction is usually restricted to the dominant pair of the group (23), with subordinate individuals engaging in a number of helping behaviors, such as territorial and nest defense, daytime incubation, and feeding of the offspring during the nestling and postfledgling stages (22). Individuals of the cohesive foraging group spend most of the time on the ground searching for invertebrates hidden in the substrate, which they excavate using their bill (22, 24). Consequently, most of the time, pied babblers forage in a head-down position within and around forbs and shrubs and hence, rely heavily on vocalizations to keep track of changes in their surroundings (21, 25–29). As such, the pied babbler vocal system exhibits around 17 discrete vocalizations, including alarm calls and sentinel calls, as well as a diverse array of social calls produced during intra- and intergroup contexts (21, 25–29).

Observational work has indicated that pied babblers produce broadband, noisy alert calls in response to sudden but generally low-urgency threats (e.g., abruptly approaching animals) and more tonal, repetitive recruitment calls when recruiting group members to a new location or during locomotion, mainly in foraging or roosting contexts. Moreover, alert and recruitment calls can be combined into a sequence on encountering and mobbing, mainly terrestrial, predators (Fig. 1). Given the context in which the two independent calls are produced, we aimed to investigate whether the sequence might, therefore, function specifically to recruit group members in a dangerous situation (e.g., when mobbing a predator) by combining information on both the danger and the requested action. Accordingly, the combination of alert and recruitment calls (hereafter termed the "mobbing sequence") might constitute a rudimentary compositional structure, where the meaning of the whole is a product of the meaning of its parts (30).

To verify the context-specific information conveyed by the independent vocalizations and test whether pied babblers extract the meaning of the sequence in a compositional way, we conducted additional natural observations combined with acoustic analyses and experimental manipulations. First, acoustic analyses were applied to confirm that alert and recruitment calls constitute two distinct vocalizations. Second, to determine the contexts in which the individual calls and the call sequence are produced, we conducted natural observations and predator presentation experiments combined with audio recordings. Third, we carried out systematic natural, artificial, and control playback experiments to investigate whether birds perceive the sequence compositionally. Key support for compositionality requires that the context in which mobbing sequences are produced and the responses of receivers to playbacks of these sequences are related to the information encoded in alert and recruitment calls (30, 31).

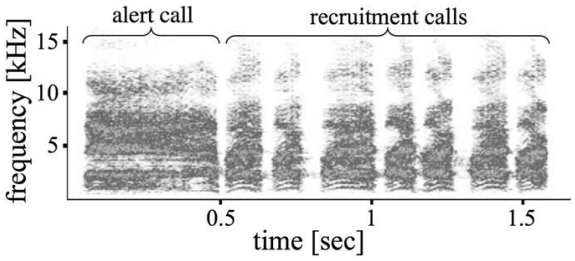


Fig. 1. Spectrogram of a mobbing sequence composed of one alert and seven recruitment calls.

Results

Acoustic Analysis. A discriminant function analysis (DFA) indicated that alert and recruitment calls could be statistically discriminated based on their structure alone ($N_{individuals} = 16$; $N_{calls} = 32$; correct classification: 97%; $P < 0.001$). When applying a leave one out cross-validated DFA, 94% were correctly classified, a classification higher than expected by chance (two-tailed binomial test; change level = 50%; $P < 0.001$).

Alert and Recruitment Calls: Natural Context. Natural observations combined with acoustic recordings were conducted to quantify the calls' context specificity. From a total of 36 alert calls recorded in 11 groups, 69% were elicited by suddenly appearing, nondangerous subjects (e.g., hares, antelopes, and researchers); 14% of alert calls were caused by inactive snakes or distant mongooses or foxes that did not present a direct threat to babblers. Another 6% were produced in response to alarm calls of con- or heterospecifics. For the remaining 11% of occasions, no obvious threat could be detected.

From a total of 196 recorded recruitment call events from 71 individuals in 20 groups, 60% resulted in other group members approaching the caller, and 6% resulted in overall group movement following the caller. In the remaining 34%, recipients either showed no response (44 of 67 occasions) or countercalled with recruitment or other loud calls (23 of 67 occasions) (29). All recorded recruitment calls were produced in nondangerous contexts in the absence of any predators. Thus, alert calls seem to encode information about low-urgency threats in a caller's imminent surrounding, and recruitment calls seem to function to recruit group members to a caller's current location.

Mobbing Sequences: Natural Context and Experimental Elicitation. We observed naturally elicited mobbing sequences on 39 occasions in 14 groups: 85% were produced in response to moving terrestrial predators (mongooses, snakes, or foxes), and 8% were produced in response to small perched raptors [pygmy falcon (*Polihierax semitorquatu*) and pearl spotted owl (*Glaucidium perlatum*)], which are assumed to only pose a threat to young, inexperienced babblers. In the remaining 8% of events, no clear context could be assigned. To experimentally confirm the context accompanying the production of mobbing sequences, babbler groups were presented with a model of a Cape cobra (*Naja nivea*), and their calling behavior was noted. From a total of 13 presentations in 10 groups, mobbing sequences were elicited 92% of the time.

Playback Experiment. To investigate the responses to mobbing sequences and their individual calls, we played back natural mobbing sequences as well as the constituent alert and recruitment calls to subjects. To rule out alternative explanations associated with the saliency of the stimulus (two vs. one call type) or priming effects (any call type preceding recruitment calls generates the same response), we implemented an additional important control condition, where we artificially replaced the alert call of a mobbing sequence with another acoustically distinct broadband babbler vocalization: the foraging "chuck" call (chuck recruitment sequence) (*Supporting Information*) (24, 32, 33). Finally, in line with previous studies (12, 13), to ensure that the key dimension for receivers was the combination of information and not any urgency-based acoustic variation encoded across the structure, as an additional control, artificial mobbing sequences

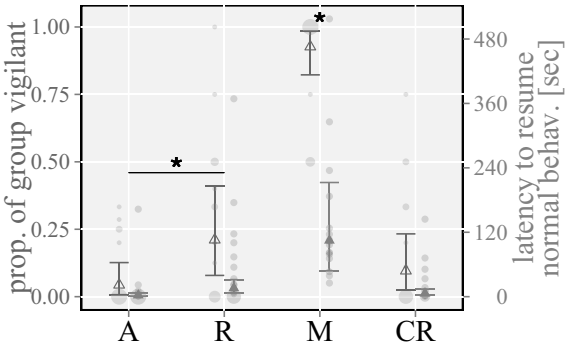


Fig. 2. Proportion of group vigilant (gray) and latency to resume normal behavior of the first reacting bird (red). Playback treatments: A, alert calls; CR, chuck recruitment sequences; M, mobbing sequences; R, recruitment calls. Bars illustrate the 95% confidence intervals, and points show the median over 16 groups of the back-transformed data. Pale circles show the raw data. *Significant differences according to the 95% confidence intervals of the difference.

were constructed from the independent calls and played back (*Supporting Information* and Table S1) (13, 32).

Our playbacks revealed differences in group attentiveness responses to the four playback conditions determined by the proportion of the group that became vigilant (treatment: $\chi^2 = 53.5$; $P < 0.01$; $n = 64$; 16 groups) (Fig. 2 and Table 1) and the latency to resume normal, nonvigilant behavior of the first reacting group member (treatment: $\chi^2 = 36.3$; $P < 0.001$; $n = 64$; 16 groups) (Fig. 2 and Table 1). Moreover, the movement patterns of a group relative to the sound source differed in response to the four stimuli (treatment: $\chi^2 = 97.2$; time: $\chi^2 = 34.9$; treatment × time: $\chi^2 = 23.6$; all $P < 0.001$; $n = 378$; 16 groups) (Fig. 3 and Table 1).

Specifically, alert calls played back in isolation did not result in noticeable changes in behavior (such as attentiveness) (Fig. 2, Table 1, and Table S2), and we found no effect of time on distance moved, with groups neither approaching nor retreating from the sound source (Fig. 3, Table 1, and Table S2). In accordance with the assumed function to recruit group members to a caller's location, in response to played back recruitment calls, babblers increased their attentiveness compared with playbacks of alert calls, likely as a way to locate the simulated recruiting caller, and slowly and steadily approached the sound source (Figs. 1 and 2, Table 1, and Table S2). Furthermore, in line with our central prediction of mobbing sequences functioning to recruit group members in a dangerous situation, we found that subjects responded most strongly to playbacks of mobbing sequences, revealing the highest attentiveness and fastest approach toward the sound source (Figs. 2 and 3, Table 1, and Table S2). Ruling out priming or stimulus effects, playbacks of chuck recruitment control sequences did not elicit similar mobbing-like behaviors, with babblers neither approaching the sound source nor increasing their attentiveness compared with playbacks of mobbing sequences (Figs. 2 and 3, Table 1, and Table S2). These results support our hypothesis that the call sequence tested conforms to the definition of basic compositional syntax (1, 14), with the high vigilance response to mobbing sequences and the fast approach to the loudspeaker being directly related to the contextual information and function of both individual calls.

Discussion

Here, we provide key comparative data indicating that the cooperatively breeding pied babbler can extract rudimentary compositional information from combinations of acoustically distinct, context-specific vocalizations: alert and recruitment calls.

Systematic observational and experimental data implementing both natural and artificial playback experiments show that pied babbler alert calls encode information on existing or imminent low-urgency threats in the environment, whereas recruitment calls communicate the motivation to recruit group members to the caller's location. Combinations of these alert and recruitments calls, here called mobbing sequences, are produced when babblers encounter and mob predominantly terrestrial threats. In response to played back mobbing sequences, babblers reacted with an increased attentiveness (high proportion of the group being vigilant and long latency to resume nonvigilant behavior) and a rapid approach toward the sound source, potentially to support the simulated caller opposing the putative threat. The context accompanying the mobbing sequence, and particularly the responses to the playbacks, suggest that the information encoded in the combination is a direct product of the constituent calls (30). We are confident that we can rule out alternative explanations related to a sequential or additive processing of calls, because responses to played back mobbing sequences exceeded those elicited by the independent calls or their sum (33, 34). Furthermore, control experiments showed that potential superstimuli (two calls vs. one call) or simple priming effects that could otherwise explain the results can be excluded, because control sequences failed to elicit similar mobbing-like behavior (32, 33). In summary, our natural observations combined with the experimental manipulations indicate that babblers produce

Table 1. Effect of playback treatments on the three response variables

Model	AICc	df	Fixed effects	χ^2	*P* value	*N*
Proportion of group vigilant						
Basic = best model: treatment	156.9	6	Treatment	53.5	<0.01	64
Intercept model	203.3	3				
Latency to resume normal behavior						
Basic = best model: treatment	242.9	6	Treatment	36.3	<0.001	64
Intercept model	272.2	3				
Movement behavior						
Basic = best model: treatment + time + treatment × time	149.0	10	Treatment	97.2	<0.001	
			Time	34.9	<0.001	378
			Treatment × time	23.6	<0.001	
Intercept model	241.2	3				

Basic models included all fixed and random effects, intercept models included only the random effects, and best models included only the significant fixed effects based on AICc selection as well as the random effects. Listed fixed effects represent the test statistics of the significant factors according to the model selection. Playback experiments were conducted on 16 groups, with each group receiving all playback treatments.

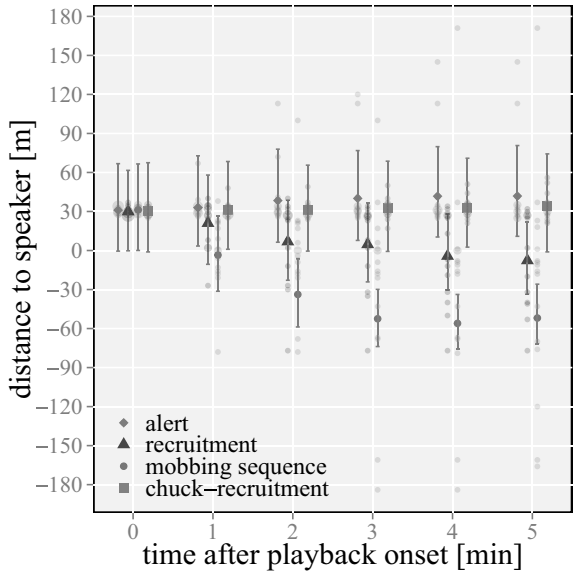


Fig. 3. Group's distance to the loudspeaker at the beginning and 1, 2, 3, 4, and 5 min after the playback start. Values over 30 m indicate a retreat from the loudspeaker. Negative values indicate that a group had passed the loudspeaker and continued moving in the same direction from where they originally heard the playback stimuli. Bars illustrate the 95% confidence intervals, and points show the median over 16 groups of the back-transformed data. Pale circles show the raw data.

and parse the sequence by linking information on the context (threat) and the requested action.

Our work, providing strong evidence for rudimentary compositionality in pied babblers, complements and extends previous research showing similar semantic combinations in primates and suggests that the basic capacity to combine meaningful calls into systematic higher-order structures may be more diverse and widespread than previously thought (8, 10, 14). Furthermore, these findings have important implications for understanding the evolutionary progression of human language. One dominant hypothesis posits that language's hierarchical syntactic system could have only evolved as part of a sudden evolutionary event, precluding the existence of intermediate protosyntactic forms (35). Alternatively, it has been suggested that syntax can be decomposed into more primitive layers, consisting of loose two- or few-word compounds that form the evolutionary and structural bases of syntactic systems (30, 36–38). Under this scenario, a sudden evolutionary leap is not necessary (30), because instead, language's syntactic complexity is hypothesized to have originally emerged out of simple but communicatively meaningful compositions. Support for this hypothesis can be found in language acquisition and newly emerging sign languages, where syntactic development initiates with simple two-word/sign compositions or "packages" (30, 38, 39) and gradually proceeds, in later stages, to more sophisticated multipackage compositions (30). Through providing comparative data for such two-signal constructs in the pied babbler vocal repertoire, our work contributes additional evidence that basic, intermediate compositional structures are viable, and hence, it supports the idea that syntax could have evolved by progressing gradually over time rather than spontaneously as an "all-or-nothing" package (37).

Exactly what evolutionary forces accompanied the progression of syntax remain elusive. Theoretical work conducted over the last two decades has aimed to disentangle the selective conditions promoting the emergence of syntax (6, 40, 41). Specifically, mathematical modeling approaches have indicated that natural selection will favor a transition toward a syntactic communication system (from a nonsyntactic one) when the number of relevant events to be communicated exceeds the number of available calls (because of either production or perception constraints) (6, 41). Our work provides relevant empirical evidence that supports this claim. Given the pied babblers' constrained vocal repertoire paired with the extensive number of social and ecological contexts that require communication (22), compositional production and processing of vocalizations are likely adaptive for pied babblers, allowing them to coordinate key additional events than would be possible with a nonsyntactic system. Moreover, combining and processing signals in a compositional way may be cognitively less demanding than evolving and memorizing new signals (41) through, for example, reinforcement learning, on the condition that the informational aspects encoded in the signals are compatible with each other. Additional experimental work, particularly natural and artificial playbacks of combinatorial and compositional structures, including temporal manipulations of the two call types, will help to shed additional light on the cognitive mechanisms involved in the parsing of call sequences.

Ultimately, however, language's generativity is not solely concerned with syntactic constructions but also, the flexible and productive concatenation of meaningful signals (40). Distinct signals or words can, for example, reoccur freely in various syntactic constructs, in a myriad of ways, and when doing so, they retain their meaning, resulting in signal compounds with overlapping or similar meaning. Although here, we show evidence for one compound signal, preliminary data suggest that babblers also flexibly combine recruitment calls with at least two additional, functionally distinct call types. Other than alert calls, recruitment calls seem to be systematically combined with aerial alarm calls when mobbing large raptors or begging calls by dependent offspring when accompanying foraging helpers (Fig. S1). These preliminary data tentatively suggest that, rather than just memorizing a complex signal, pied babblers may apply a general combinatorial rule to encode multiple messages.

In conclusion, our work provides evidence for rudimentary compositional syntax in a social bird. We propose that, through studying highly social species with discrete, constrained vocal repertoires, additional light can be shed on the variation and distribution of combinatorial mechanisms outside of humans. We predict that comparative work will, in turn, help elucidate the evolutionary drivers promoting the emergence of syntactic communication in animals and ultimately, humans.

Materials and Methods

Study Site and Species. The study was conducted on a population of wild, free living southern pied babblers at the Pied Babbler Research Project, Kuruman River Reserve in the Kalahari Desert of South Africa (26°58′S, 21°49′E). The study site is characterized by sparse vegetation and a semiarid climate (42). The population is part of a long-term research project founded by A.R.R. in 2003. Individuals are habituated to human observers and can be followed at a distance of 1–2 m, enabling close observations (24). Colored rings allow individual identification of all members of the study population (24).

General Information. Natural observations were conducted from January to April/May 2014 and 2015. The rest of the study was performed between February and April, 2014. All audio recordings were conducted using a Rode NTG-2 Directional Microphone (sampling frequency of 48 kHz; 24-bits accuracy) coupled with a Rode Blimp Suspension Windshield (Rode Microphones) and a Roland R-26 Portable Recorder (Roland Corporation). The study was performed under the permission of the ethical committee for animal research of the University of Cape Town and the Northern Cape Conservation Authority, South Africa.

Acoustic Analysis. To verify that mobbing sequences are composed of two structurally distinct call types (i.e., alert and recruitment calls), we conducted basic acoustic analyses. To avoid erroneous *P* value estimation associated

with pseudoreplication, we only took 1 alert and 1 recruitment call per individual per group, totaling 32 calls from 16 different individuals belonging to 16 different groups (43). Calls were initially inspected and assessed for quality (signal to noise ratio), and both calls that were produced as part of a sequence as well as calls produced in isolation were included in the analyses. Because alert calls sometimes lacked a clear fundamental frequency, calls were compared based on parameters related more to time and energy distribution. Additionally, we assessed the percentage of the call that exhibited clear, tonal structures (i.e., did not exhibit noise or deterministic chaos) (44). The following acoustic measurements were recorded: call duration; 25%, 50%, and 75% energy quartiles; relative time of maximum intensity; amplitude variation; amplitude rate; shimmer; and percentage of voiced structures in the first and second halves of the call. Except the latter, all parameters were extracted using an automated, custom-built analysis script in Praat 5.1.03. To determine the classification probabilities of calls to call type (alert or recruitment), we first applied a DFA using SPSS (version 21.0; IBM). Only parameters with a variance inflation factor smaller than 10 were included in the analysis. Depending on the number of groups to be classified, the DFA creates one or more discriminant functions by identifying linear combinations of the predictor variables that best describe the discrimination between groups (45). A leave one out cross-validation procedure was applied for external validation. A two-tailed binomial test was used to estimate the overall significance of the classification of the DFA, with a corrected level of chance corresponding to the number of categories discriminated (two categories = 50%).

Natural Observations. To quantify the context in which alert calls, recruitment calls, and mobbing sequences are produced, natural observations combined with audio recordings were conducted. In 2014, we regularly visited 19 babbler groups with an average group size of 6.2 ± 2.3 individuals, and in 2015, we regularly visited 18 groups with an average group size of 5.1 ± 1.4 individuals. A specific group was followed in the evening for approximately 2 h until the group had settled down in a night roost. The next morning, the group was rejoined at the sleeping roost before dawn and then followed for around 4 h. Whole sessions were audio recorded, and they were annotated and analyzed using Cool Edit 2000 (Syntrillium Software Corporation) or Audition CS6 (Adobe), scanning for relevant events and vocalizations.

Predator Presentation Experiment. Presentation experiments were conducted to verify the context-specific production of mobbing sequences when mobbing predators. Ten babbler groups were exposed to a rubber snake simulating an active Cape cobra with an extended neck. The model was placed in a raised posture along the predicted path of a group either below vegetation or coiled around the trunk of a tree. The whole procedure was audio recorded and analyzed using Audition CS6 (Adobe) to determine whether mobbing sequences (i.e., combinations of alert and recruitment calls) were produced.

Playback Stimuli. For the creation of playback stimuli, high signal to noise ratio vocalizations of male or female subordinate group members from each of the test groups were selected. However, in one group only vocalizations from a dominant individual could be recorded. Playbacks were created and normalized with Audition CS6 (Adobe; sampling frequency of 48 kHz; 24-bits accuracy). To test whether the mobbing sequence derives its meaning from the meaning of its individual calls and verify behavioral observations suggesting context-specific production of the constituent calls, we played back natural mobbing sequences as well as alert calls and recruitment calls on their own to subjects. Because the response to the mobbing sequence could have been the result of simple priming or stimulus intensity effects, any acoustic element preceding recruitment calls or equally, any two call types in combination could have been sufficient to elicit the behavioral change (32, 33). To exclude these possibilities, we created a two-call control chuck recruitment sequence (*Supporting Information*). This control combination was created by replacing the alert call of the mobbing sequence with a chuck call [contact/close call produced during foraging (24)] of the same individual. The chuck call was, therefore, normalized to the amplitude of the substituted alert call, and the same interelement distance between the replaced element and the recruitment call was maintained.

To rule out that any urgency-based acoustic information encoded in the naturally occurring sequence might have elicited a mobbing-like response, we created two sets of stimuli versions for the playback experiments. The first set included natural mobbing sequences, the constituent alert and recruitment calls that were played back in isolation, as well as the chuck recruitment sequence created out of the natural mobbing sequence. The second set included artificially created mobbing sequences created by synthetically combining alert and recruitment calls (*Supporting Information*).

Playback Protocol and Response Variables. Stimuli were played back once at a naturally occurring, normalized amplitude (~73 dB at a 4 m distance) using an AN-30 Speaker Monitor (Anchor) coupled to an iPod 3 (Apple Inc.). Sound files were uploaded to the iPod which was remote controlled via Bluetooth using an iPhone 4 (Apple Inc.) and Tango Remote App (Blue Atlas Technology, LLC). Each of the 16 test groups was exposed to all four playback conditions in a randomized order, and only vocalizations of an existing group member were played back. All four treatment conditions were played back in one morning, except for in one occasion, when one condition had to be played back on a separate day because of experiment interruption by a predator. The loudspeaker was placed ~30 m from the target group and hidden by vegetation. Playbacks were conducted when no individual was on sentinel duty and when no major disturbances had occurred on the morning that the playbacks were undertaken. In line with our prediction of mobbing sequences functioning to recruit group members in a dangerous situation, we recorded subjects' vigilance responses as well as movement patterns. After the playback started, the proportion of individuals that became vigilant was recorded. Vigilance was classified as scanning the area or looking toward the location from where the stimulus was broadcast. To avoid including individuals that simply became attentive in response to an alert group member, only individuals that reacted immediately after the stimulus presentation were counted as vigilant. Additionally, the latency for the first responding bird to resume normal (nonvigilant) behavior was recorded. To evaluate differences in movement behavior (direction and speed), the distance from the spatial center of the group to the loudspeaker was recorded at the beginning of the playback and after 1, 2, 3, 4, and 5 min using a handheld global positioning system-logger (Garmin eTrex 10; Garmin Ltd.) and Garmin Basecamp software (Garmin Ltd.). All experiments were videotaped using a Sony Handycam (HDR-CX160). Videos were analyzed frame by frame using Audition CS6 (Adobe).

Statistical Analysis of Playback Experiment. Statistical analyses were conducted in R (version 3.1.1) (46). For the computation of linear mixed models (LMMs) and generalized LMMs, the packages lme4 (47) and MuMIn (48) were used. Model estimates were plotted using the packages ggplot2 (49) and gtable (50). Model selection was based on Akaike's information criterion corrected for small sample sizes (AICc), with a threshold difference (ΔAICc) of at least two to the next best model (51). If the difference between the model with the lowest AICc and subsequent models was less than two, the influence of each fixed factor on the response variable was assessed for each of the models within the specified range. The best model was then chosen by excluding the model(s) that included nonsignificant predictor variables. The significance of the fixed effects was assessed based on bootstrapping methods. Therefore, data were simulated based on the null model (best model according to model selection excluding the factor of interest). The full model (best model according to model selection) and the null model were then fitted to the simulated data, and their difference in deviance was calculated. Simulations and model fittings were iterated 10,000 times. The same procedure was repeated, but in this instance, the actual data was fit to the null and full models. The distributions of differences in deviances obtained with the simulated data and the actual data were then compared by applying a χ^2 test [see also R pbkrtest package (52)]. To investigate where the differences between the playback conditions are, the 95% confidence intervals of the difference were compared between each treatment conditions. If the confidence intervals intersected zero, differences were nonsignificant (51).

Model 1: Proportion of group vigilant. To test for an effect of the playback type, we fitted a generalized LMM with a binomial error distribution (0–1 = proportion of group vigilant), with the number of vigilant individuals representing the response term and group size representing the binomial denominator. Because of a possible zero inflation, overdispersion in the model was estimated by counting each variance parameter as one df. The data were considered overdispersed if the ratio of the sum of squared Pearson residuals to residual dfs was greater than one, which was true in our model (53). To correct for overdispersion, an observational-level random term was added to the model by serially numbering each observation (54). Accordingly, model 1 included the treatment type as a fixed effect and the group identity and the observation level as random effects.

Model 2: Latency to resume normal behavior. To examine whether the playback condition had an effect on the latency to resume normal, nonvigilant behavior of the first reacting bird, we fitted an LMM with treatment type as a fixed effect and group identity as a random effect. To achieve a normal distribution, the data were log-transformed.

Model 3: Movement behavior. To investigate differences in movement behavior over time between the playback conditions, a group's distance to the sound source was recorded at fixed time intervals. After a group had passed the loudspeaker and continued moving in the direction from where they

originally heard the stimuli, negative values for the distance to the speaker were assigned. An LMM was fitted with treatment type, time, and its interaction term as fixed effects and group identity as a random effect. To achieve a normal distribution, the data were log-transformed, with a constant value being added to the response variable to avoid transformation of negative values [i.e., $\log(x + 200)$] (55).

ACKNOWLEDGMENTS. We thank T. Clutton-Brock, M. Manser, D. and N. Gaynor, the Kalahari Research Trust, Mr. and Mrs. de Bruin, and Mr. and Mrs. Kotze for access to land and logistics; E. Wiley, J. Westrip, R. Mendonça, R. Hopper, A. Thompson, D. Humphries, F. Finch, M. Nelson-Flower, and T. Flower for maintaining habituation of the study population, collecting life history data, and discussions; E. Mandel-Briefer for help with the acoustic analysis and providing the custom Praat script for the acoustic analysis; M. Ferrari, A. Sutter, and the IEU R lunch group for statistical advice; K. Messenger, K. Sotaro, and B. Bickel for discussions; and A. Russell, K. Collier, and three anonymous reviewers for constructive comments on previous versions of the manuscript. Particular thanks to M. Manser for input, discussions, and support. Funding was provided by Forschungskredit of the University of Zurich Grants 57191601 (to S.E.) and FK-14-077 (to S.E.), Swiss National Science Foundation Grants P1ZHP3_151648 (to S.E.) and 31003A_153065 (to S.W.T.), and the Claraz Stiftung (to S.W.T.).

1. Hurford J (2011) Linguistics from an evolutionary point of view. Handbook of the Philosophy of Science: Linguistics, eds Kempson R, Fernando M, Ashe N (Elsevier Science, Amsterdam), Vol 14, pp 473–498.
2. Chomsky N (1981) Knowledge of language: Its elements and origins. *Philos Trans R Soc Lond B Biol Sci* 295(1077):223–234.
3. Humboldt Wv, *On Language*, trans Heath P (1999) (Cambridge Univ Press, Cambridge, United Kingdom).
4. Hauser MD, Chomsky N, Fitch WT (2002) The faculty of language: What is it, who has it, and how did it evolve? *Science* 298(5598):1569–1579.
5. Christiansen MH, Kirby S (2003) Language evolution: Consensus and controversies. *Trends Cogn Sci* 7(7):300–307.
6. Nowak MA, Plotkin JB, Jansen VAA (2000) The evolution of syntactic communication. *Nature* 404(6777):495–498.
7. Harvey PH, Pagel MD (1991) *The Comparative Method in Evolutionary Biology* (Oxford Univ Press, Oxford).
8. Ouattara K, Lemasson A, Zuberbühler K (2009) Campbell's monkeys use affixation to alter call meaning. *PLoS One* 4(11):e7808.
9. Arnold K, Zuberbühler K (2008) Meaningful call combinations in a non-human primate. *Curr Biol* 18(5):R202–R203.
10. Arnold K, Zuberbühler K (2012) Call combinations in monkeys: Compositional or idiomatic expressions? *Brain Lang* 120(3):303–309.
11. Schlenker P, et al. (2014) Monkey semantics: Two 'dialects' of Campbell's monkey alarm calls. *Linguist and Philos* 37(6):439–501.
12. Coye C, Ouattara K, Zuberbühler K, Lemasson A (2015) Suffixation influences receivers' behaviour in non-human primates. *Proc Bio Sci* 282(1807):20150265.
13. Arnold K, Zuberbühler K (2006) Language evolution: Semantic combinations in primate calls. *Nature* 441(7091):303.
14. Collier K, Bickel B, van Schaik CP, Manser MB, Townsend SW (2014) Language evolution: Syntax before phonology? *Proc Biol Sci* 281(1788):20140263.
15. Zuberbühler K (2015) Linguistic capacity of non-human animals. *Wiley Interdiscip Rev Cogn Sci* 6(3):313–321.
16. Catchpole CK, Slater TLB (1995) *Bird Song: Themes and Variations* (Cambridge Univ Press, New York).
17. Berwick RC, Okanoya K, Beckers GJ, Bolhuis JJ (2011) Songs to syntax: The linguistics of birdsong. *Trends Cogn Sci* 15(3):113–121.
18. Rendall D (2013) Q&A: Cognitive ethology—inside the minds of other species. *BMC Biol* 11(108):108.
19. Suzuki TN, Wheatcroft D, Griesser M (2016) Experimental evidence for compositional syntax in bird calls. *Nat Commun* 7:10986.
20. Raihani NJ (2008) Cooperation and conflict in pied babblers. PhD thesis (University of Cambridge, Cambridge, United Kingdom).
21. Golabek KA (2010) Vocal communication and the facilitation of social behaviour in the southern pied babbler (Turdoides bicolor). PhD thesis (University of Bristol, Bristol, United Kingdom).
22. Ridley AR, Raihani NJ (2007) Variable postfledging care in a cooperative bird: Causes and consequences. *Behav Ecol* 18(6):994–1000.
23. Nelson-Flower MJ, et al. (2011) Monogamous dominant pairs monopolize reproduction in the cooperatively breeding pied babbler. *Behav Ecol* 22(3):559–565.
24. Radford AN, Ridley AR (2008) Close calling regulates spacing between foraging competitors in the group-living pied babbler. *Anim Behav* 75(2):519–527.
25. Bell MB, Radford AN, Rose R, Wade HM, Ridley AR (2009) The value of constant surveillance in a risky environment. *Proc Biol Sci* 276(1669):2997–3005.
26. Hollén LI, Bell MBV, Radford AN (2008) Cooperative sentinel calling? Foragers gain increased biomass intake. *Curr Biol* 18(8):576–579.
27. Radford AN, Ridley AR (2006) Recruitment calling: A novel form of extended parental care in an altricial species. *Curr Biol* 16(17):1700–1704.
28. Radford AN, Ridley AR (2007) Individuals in foraging groups may use vocal cues when assessing their need for anti-predator vigilance. *Biol Lett* 3(3):249–252.
29. Golabek KA, Radford AN (2013) Chorus-call classification in the southern pied babbler: Multiple call types given in overlapping contexts. *Behaviour* 150(7):691–712.
30. Hurford J (2011) *The Origins of Grammar* (Oxford Univ Press, Oxford).
31. Coye C, Zuberbühler K, Lemasson A (2016) Morphologically structured vocalizations in female Diana monkeys. *Anim Behav* 115:97–105.
32. Engesser S, Crane JM, Savage JL, Russell AF, Townsend SW (2015) Experimental evidence for phonemic contrasts in a nonhuman vocal system. *PLoS Biol* 13(6):e1002171.
33. Slocombe KE, Kaller T, Call J, Zuberbühler K (2010) Chimpanzees extract social information from agonistic screams. *PLoS One* 5(7):e11473.
34. Scott-Phillips TC, Gurney J, Ivens A, Diggle SP, Popat R (2014) Combinatorial communication in bacteria: Implications for the origins of linguistic generativity. *PLoS One* 9(4):e95929.
35. Berwick RC (2011) All you need is merge: Biology, computation, and language from the bottom-up. *The Biolinguistic Enterprise*, eds Sciullo AMD, Boeckx C (MIT Press, Cambridge, MA), pp 706–825.
36. Jackendoff R (1999) Possible stages in the evolution of the language capacity. *Trends Cogn Sci* 3(7):272–279.
37. Progovac L (2015) *Evolutionary Syntax* (Oxford Univ Press, Oxford).
38. Tomasello M (2003) *Constructing a Language: A Usage-Based Theory of Language Acquisition* (Harvard Univ Press, Cambridge, MA).
39. Senghas A, Kita S, Ozyürek A (2004) Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science* 305(5691): 1779–1782.
40. Nowak MA, Komarova NL (2001) Towards an evolutionary theory of language. *Trends Cogn Sci* 5(7):288–295.
41. Nowak MA, Krakauer DC (1999) The evolution of language. *Proc Natl Acad Sci USA* 96(14):8028–8033.
42. Clutton-Brock TH, et al. (1998) Costs of cooperative behaviour in suricates (Suricata suricatta). *Proc Biol Sci* 265(1392):185–190.
43. Sharp SP, McGowan A, Wood MJ, Hatchwell BJ (2005) Learned kin recognition cues in a social bird. *Nature* 434(7037):1127–1130.
44. Fitch WT, Neubauer J, Herzel H (2002) Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Anim Behav* 63(3):407–418.
45. Mundry R, Sommer C (2007) Discriminant function analysis with nonindependent data: Consequences and an alternative. *Anim Behav* 74(4):965–976.
46. R Core Team (2014) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).
47. Bates D, Maechler M, Bolker B, Walker S (2014) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67(1):1–48.
48. Bartón K (2014) MuMIn: Multi-model inference. R package. Available at CRAN.R-project.org/package=MuMIn. Version 1.10.5.
49. Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York).
50. Wickham H (2012) gtable: Arrange grobs in tables. R package. Available at CRAN.R-project.org/package=gtable. Version 0.1.2.
51. Burnham KP, Anderson DR (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer, New York), 2nd Ed.
52. Halekoh U, Højsgaard S (2014) A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models - the R package pbkrtest. *J Stat Softw* 59(9):1–30.
53. Zuur A, Ieno EN, Walker N, Saveliev AA, Smith GM (2009) *Mixed Effects Models and Extensions in Ecology with R* (Springer, New York).
54. Harrison XA (2014) Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ* 2:e616.
55. LeBlanc DC (2004) *Statistics: Concepts and Applications for Science* (Jones and Bartlett Publishers, Mississauge, ON, Canada).

Supporting Information

Engesser et al. 10.1073/pnas.1600970113

Stimuli Sets

Stimuli Sets for Playback Experiments. To verify that pied babblers (*Turdoides bicolor*) perceive the sequence as a combination of two call types and rule out that any urgency-based acoustic information encoded in the sequence might have resulted in a mobbing-like response, we played back artificially constructed mobbing sequences composed of alert calls combined with recruitment calls originating from group travel events (instead of mobbing events). This control further tested whether recruitment calls produced in (dangerous) mobbing and (nondangerous) group travel contexts are acoustically identical and also perceived as such by receivers. Playbacks were conducted on 16 groups, with each group being exposed to all four playback conditions (alert calls, recruitment calls, mobbing sequence, and chuck recruitment sequence). Of 16 tested groups, 8 groups received a stimuli set originating from original mobbing events (stimuli set A). For the remaining eight groups, an artificial stimuli set was created (stimuli set B), totaling 16 unique, group-specific stimuli sets.

Stimuli set A. Original mobbing sequences composed of alert and recruitment calls were obtained from either natural mobbing events or mobbing events induced by presenting a Cape cobra (*Naja nivea*) model to babbler groups. The calls for the alert and recruitment call treatments were directly extracted from the mobbing sequence. For the chuck recruitment sequence, the alert call in the mobbing sequence was replaced by a chuck call of the same individual. Hence, both recruitment calls and chuck recruitment sequences were composed of recruitment calls originating from a mobbing event.

Stimuli set B. Artificial mobbing sequences were constructed by combining an alert call with recruitment calls that were produced during group travel, both originating from the same individual. The recruitment calls used for the recruitment and chuck recruitment conditions for those stimuli in stimuli set B were the same as those used in the artificial mobbing sequence (i.e., originated from the specific group travel event). To match the natural variation, artificial mobbing sequences were composed of 1–2 alert calls [two calls in cases where alert calls were particularly short in duration (two instances)] and 4–7 recruitment calls.

Statistical Analysis and Results of the Effect of the Stimuli Set Type. In 8 of 16 groups, a stimuli set originating from original mobbing events (stimuli set A) was played back, and the remaining 8 groups received an artificial stimuli set (stimuli set B). Accordingly, the recruitment calls used in the two sets differed, with the ones of stimuli set A originating from (dangerous) mobbing events and the ones from stimuli set B originating from (nondangerous) group travel events. Recruitment calls were, however, only presented in three of four playback conditions (i.e., the treatments recruitment call, mobbing sequence, and chuck recruitment sequence but not the alert call treatment). Accordingly, in a first step, we examined whether the version of recruitment calls (version A of stimuli set A vs. version B of stimuli set B) had an effect on our response variables by only including those three conditions in our analyses (see Table S2). Model selection based on the AICc values resulted in the recruitment call version being excluded as a factor from all models, indicating that whether recruitment calls originated from mobbing or group travel events did not influence the proportion of vigilant group members, the latency to resume normal behavior, or the movement behavior of the group (Table S1). Hence, babblers responded the same way to natural and artificial mobbing sequences, recruitment calls extracted from mobbing sequences and recruitment calls produced in group travel context, and chuck recruitment sequences constituting recruitment calls from the two opposing events. Based on the lack of an effect of recruitment call version, we proceeded with the analyses of the full experiment, including all four playback conditions (in the text).

Acoustic Analysis of Chuck, Alert, and Recruitment Calls. To rule out priming or stimulus intensity effects of the mobbing sequence, we created an artificial control sequence, where we replaced the alert call of the mobbing sequence with a babbler chuck call. To assess the discriminability of chuck vocalizations compared with alert and recruitment calls, we carried out an additional acoustic and discriminant function analysis (DFA), whereby we integrated chuck vocalizations into the initial alert and recruitment call acoustic dataset. To control for individual identity and avoid pseudoreplication, 16 chuck calls were used, originating from the same 16 individuals from which alert and recruitment calls were obtained. The extraction of the acoustic parameters and the DFA were conducted as described in the text (*Materials and Methods*). The 50%-energy quartile was excluded from the analysis due to a variance inflation factor greater than 10. The DFA indicated that chuck, alert, and recruitment calls could be statistically discriminated with a correct classification of 96% ($P < 0.001$; $N_{\text{individuals}} = 16$; $N_{\text{calls}} = 48$); 94% of calls were classified correctly when applying a leave one out cross-validated DFA (two-tailed binomial test; chance level = 33%; $P < 0.001$).

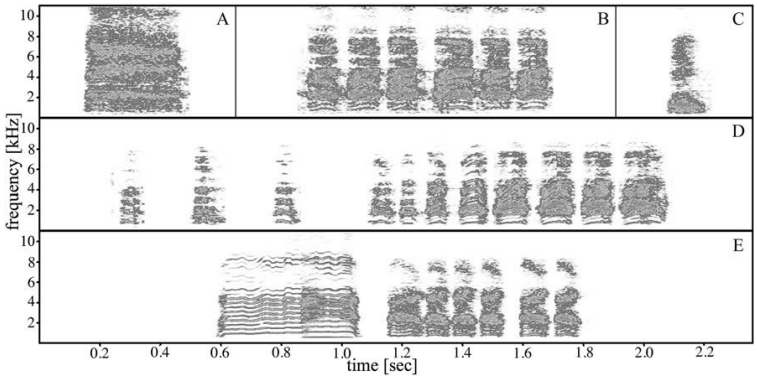


Fig. S1. Spectrograms of a babbler (*A*) alert call, (*B*) recruitment calls, and (*C*) chuck call from the same individual. (*D*) Natural combination of aerial alarm calls with recruitment calls. (*E*) Natural combination of a begging call with recruitment calls.

Table S1. Results of analyses excluding the alert call playback condition testing for a possible effect of recruitment call version

Model	AICc	df	Fixed effects	χ^2	*P* value	*N*
Proportion of group vigilant						
Basic model: treatment + version + treatment × version	132.6	8				
Intercept model	162.3	3				
Best model: treatment	129.2	5	Treatment	38.0	<0.01	48
Latency to resume normal behavior						
Basic model: treatment + version + treatment × version	187.4	8	Treatment	27.7	<0.001	48
			Version	6.5	0.13	48
			Treatment × version	5.4	0.11	48
Intercept model	202.6	3				
Best model: treatment	185.6	5	Treatment	21.9	<0.001	48
Movement behavior						
Basic model: treatment + time + version + treatment × time	181.6	9	Treatment	56.5	<0.001	282
			Time	27.5	<0.001	282
			Treatment × time	13.4	0.001	282
			Version	1.5	0.26	282
Intercept model	239.8	3				
Best model: treatment + time + treatment × time	181.0	8	Treatment	56.3	<0.001	282
			Time	27.6	<0.001	282
			Treatment × time	13.3	0.001	282

Basic models included all fixed and random effects, intercept models included only the random effects, and best models included only the significant fixed effects based on AICc selection as well as the random effects. Listed fixed effects represent the test statistics of the factors that were, according to the model selection, significant. Playback experiments were conducted on 16 groups, with each group receiving all playback treatments.

Table S2. Post-hoc analyses of playback experiment

Contrast	2.5%	97.5%
Movement behavior		
SD_(intercept)|group	0.04*	0.14*
σ	0.25*	0.30*
A	5.34*	5.56*
A vs. R	−0.15	0.13
A vs. M	−0.21	0.07
A vs. CR	−0.15	0.14
Time	−0.02	0.04
A × time vs. R × time	−0.09*	0.00*
A × time vs. M × time	−0.15*	−0.06*
A × time vs. CR × time	−0.06	0.04
SD_(intercept)|group	0.04*	0.14*
σ	0.25*	0.30*
R	5.33*	5.55*
R vs. A	−0.12	0.16
R vs. M	−0.21	0.08
R vs. CR	−0.13	0.15
Time	−0.07*	−0.01*
R × time vs. A × time	0.00*	0.09*
R × time vs. M × time	−0.10*	−0.01*
R × time vs. CR × time	−0.01	0.08
SD_(intercept)|group	0.04*	0.14*
σ	0.26*	0.30*
M	5.25*	5.47*
M vs. A	−0.05	0.22
M vs. R	−0.07	0.20
M vs. CR	−0.08	0.22
Time	−0.12*	−0.06*
M × time vs. A × time	0.06*	0.14*
M × time vs. R × time	0.01*	0.10*
M × time vs. CR × time	0.05*	0.15*
SD_(intercept)|group	0.04*	0.13*
σ	0.26*	0.30*
CR	5.34*	5.55*
CR vs. A	−0.13	0.13
CR vs. R	−0.15	0.12
CR vs. M	−0.20	0.06
Time	−0.03	0.03
CR × time vs. A × time	−0.03	0.05
CR × time vs. R × time	−0.08	0.01
CR × time vs. M × time	−0.14*	−0.05*
Proportion of group vigilant		
SD_(intercept)|observation	0.46*	2.17*
SD_(intercept)|group	0.00*	1.00*
A	−4.88*	−1.93*
A vs. R	0.30*	3.65*
A vs. M	4.06*	8.42*
A vs. CR	−0.70	2.65
SD_(intercept)|observation	0.79*	2.33*
SD_(intercept)|group	0.00*	0.99*
R	−2.46*	−0.37*
R vs. A	−3.54*	−0.25*
R vs. M	2.40*	5.87*
R vs. CR	−2.52	0.52
SD_(intercept)|observation	0.45*	2.15*
SD_(intercept)|group	0.00*	1.10*
M	1.53*	4.11*

Table S2. Cont.

Contrast	2.5%	97.5%
M vs. A	−8.47*	−3.98*
M vs. R	−6.13*	−2.36*
M vs. CR	−7.50*	−3.34*
SD_(intercept)|observation	0.00*	2.19*
SD_(intercept)|group	0.00*	1.12*
CR	−3.69*	−1.19*
CR vs. A	−2.62	0.68
CR vs. R	−0.61	2.39
CR vs. M	3.26*	7.07*
Latency to relax		
SD_(intercept)|group	0.00*	0.89*
σ	1.16*	1.75*
A	0.58*	2.08*
A vs. R	0.52*	2.51*
A vs. M	2.31*	4.28*
A vs. CR	−0.30	1.69
SD_(intercept)|group	0.00*	0.86*
σ	1.17*	1.76*
R	2.10*	3.50*
R vs. A	−2.48*	−0.53*
R vs. M	0.79*	2.87*
R vs. CR	−1.87	0.17
SD_(intercept)|group	0.00*	0.85*
σ	1.17*	1.73*
M	3.90*	5.36*
M vs. A	−4.43*	−2.37*
M vs. R	−2.93*	−0.76*
M vs. CR	−3.60*	−1.57*
SD_(intercept)|group	0.00*	0.84*
σ	1.14	1.71*
CR	1.26*	2.69*
CR vs. A	−1.62	0.30
CR vs. R	−0.12	1.96
CR vs. M	1.69*	3.69*

The 95% confidence intervals (95% CIs) of the differences were compared based on bootstrapping methods. Each treatment type was leveled separately and compared with the 95% CIs of the remaining treatments. Playback conditions: A, alert calls; CR, chuck recruitment sequence; M, mobbing sequence; R, recruitment calls.

*95% CIs not crossing zero were categorized as a significant contrast.

CHAPTER 4

This way! Pied babblers modify the internal structure of their recruitment cries specifying the form of recruitment

Submitted to Proceedings of the Royal Society B

This way! Pied babblers modify the internal structure of their recruitment cries specifying the form of recruitment

Sabrina Engesser [a], Amanda R. Ridley [b,c], Marta B. Manser [a], Simon W. Townsend [a,d]

[a] Animal Behaviour, Department of Evolutionary Biology and Environmental Studies, University of Zurich, 8057 Zurich, Switzerland

[b] Centre for Evolutionary Biology, School of Animal Biology, The University of Western Australia, Crawley WA 6009, Australia

[c] Percy FitzPatrick Institute, University of Cape Town, Rondebosch 7701, South Africa

[d] Department of Psychology, University of Warwick, Coventry CV4 7AL, United Kingdom

Abstract

Language is inherently combinatorial. Parallels can be found in non-human systems, with animals recombining 'meaningless' sounds to create functionally distinct vocalisations, or assembling 'meaningful' units into higher-order meaningful sequences. However, further analogue examples are central in determining how conserved and variable combinatorial structures are, and to identify factors promoting the emergence of generative vocal mechanisms. Here we provide evidence for a novel type of combinatoriality within the sub-structure of pied babbler (*Turdoides bicolor*) 'recruitment cries'. Using acoustic analyses, behavioural observations and playback experiments, we demonstrate that males produce two functionally distinct recruitment cry types. While both types are similar in overall super-structure, they differ in their internal structure, comprising either repetitions of single-note or double-note motifs. Consistent with similarities in super-structure, our findings indicate that both cries signal the intention to recruit the group, with the internal sub-structure specifying the form of recruitment, from requesting approach to following over longer distances. We argue this work illustrates the diversity of combinatorial mechanisms outside of human language, and supports the hypothesis that combinatoriality evolved in various species potentially to offset the limitations of small vocal repertoires and to increase communicative output.

Keywords: recruitment - animal communication - combinatoriality - acoustic modifier - language evolution - *Turdoides bicolor*

Introduction

One of language's defining feature is its expressive power. By flexibly combining a limited suite of acoustic elements, a theoretically infinite myriad of messages can be communicated [1-3]. Understanding the factors that lead to the evolution of such combinatoriality is a fundamental question in the field of language evolution [4, 5]. One dominant hypothesis derived from theoretical and computational modelling approaches suggests that combinatoriality is one evolutionary solution to tradeoff vocal constraints that limit the number of distinct sounds that can be produced and perceived [6-8]. Combining sounds can serve to create more easily distinguishable signals or words, and assembling such units into higher-order meaningful structures can further offset memory limitations and increase the amount of information that can be communicated [6-8].

Comparative research on animals provides a powerful tool to assess this hypothesis [4, 9]. Recent empirical studies have demonstrated that stringing both 'meaningless' and 'meaningful' sounds together can function to increase the communicative output of a species [10-13]. For example, chestnut-crowned babblers (*Pomatostomus ruficeps*) reuse and rearrange acoustic elements to create functionally distinct and perceptually relevant flight and provisioning calls [14]. Banded mongooses (*Mungos mungo*) and Diana monkeys (*Cercopithecus diana*) produce two-compound utterances, with one unit encoding the caller's identity and the other relating to the caller's behaviour or motivational state [15-17]. On a structurally higher level, putty-nosed monkeys (*Cercopithecus nictitans*) combine predator-specific alarm calls into a sequence that functions to initiate group movement [18-20], and Campell's monkeys (*Cercopithicus cambelli*) affix alarm calls with a meaning-modifying morpheme [21, 22]. Furthermore, two recent experimental studies in non-primates suggest these abilities are more widespread than previously thought; both pied babblers (*Turdoides bicolor*) and Japanese great tits (*Parus major minor*) have been shown to combine alert and recruitment calls when mobbing predators, potentially linking information on the danger with the requested action [23, 24]. Such data are particularly important in helping to elucidate the various forms of combinatoriality that exist in animal communication, and critically the mechanisms by which new meaning can be generated.

Here, we extend this body of data by demonstrating a novel form of combinatorial structuring in pied babbler 'recruitment cries', with internal acoustic variation within the larger cry structure appearing to refine the functional specificity of the signal. Pied babblers are cooperatively breeding passerines that possess a fixed vocal repertoire of around 17 discrete vocalisations, including alarm calls and an array of social calls [25-32]. They live in stable groups of up to 15 individuals with reproduction restricted to the dominant pair and subordinate helpers assisting in rearing the offspring [33, 34]. Members of the group cooperatively defend their territory (average territory size 0.75km^2 [35]), and forage and travel in a highly cohesive manner [36, 37]. Babblers primarily forage on the ground and are weak flyers, with flight movements between foraging or resting sites typically being short and low to the ground [36]. The high degree of sociality in this species, and the diverse behaviours that require coordination makes them an ideal candidate to investigate

combinatorial mechanisms that may have evolved to counter vocal constraints and to increase the range of events and behaviours that can be communicated.

While previous work has demonstrated that pied babblers combine alert and recruitment calls into 'mobbing sequences' to recruit group members in dangerous situations [23], recent observational work has shown that further internal structural variation characterises the recruitment component of the mobbing sequence (personal observation). Specifically, when groups move between foraging or diurnal resting sites, males (predominantly the dominant male of a group) produce two types of loud calls: 'atonal chatters' and 'double-note ascending calls' [32, 38], which we, for clarity, refer to as 'single-note recruitment cries' (SN cries) and 'double-note recruitment cries' (DN cries). Both cry types start with a wind-up element which increases in amplitude and grades into a repetition of either single-note motifs (acoustic patterns) [$(A)^n$] for SN cries, or double-note motifs [$(AB)^n$] for DN cries, with both motif types sharing the A note (the equivalent to the 'recruitment calls' of the mobbing sequence [23]), and a B note being added to each A note in DN cries (Fig. 1a, b). Anecdotal observations suggest that SN cries function to recruit group members to a caller's current location. DN cries, on the other hand, appear to initiate a cohesive group movement over a longer distance, potentially as a way to patrol the group's existing territory, with a caller vocalising prior to departure, and receivers following the caller's movement.

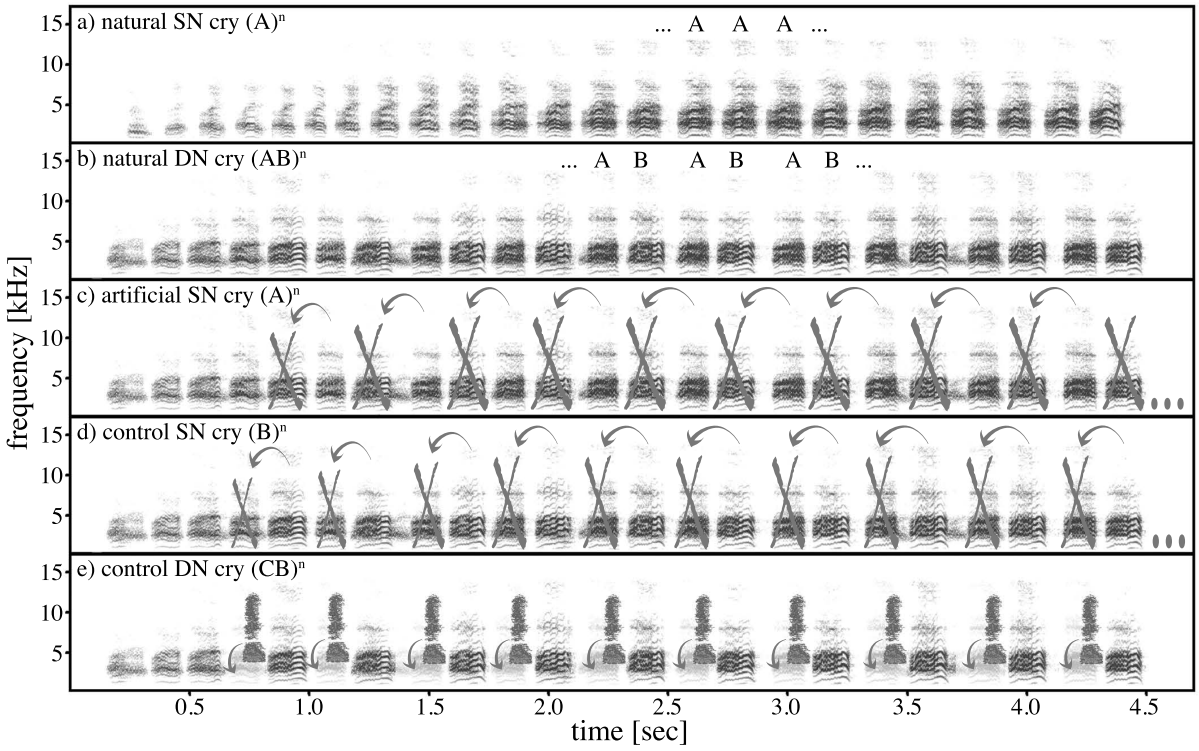


Figure 1. a, b) Spectrograms of a natural SN cry and a natural DN cry of the same dominant male. Capital letters denote the note type. c-e) Example spectrograms of manipulated cries used for playback experiments: c) artificial SN cry generated from a DN cry by deleting each B note; d) control SN cry generated from a DN cry by deleting each A note, e) control DN cry generated by substituting each A note with a chuck vocalisation.

Using acoustic analyses, behavioural observations, and playback experiments we investigated the structural similarity, contextual distinctiveness, and function of the two recruitment cry variants. Firstly, acoustic analyses served to verify whether DN cries are composed of 2 distinct note types (A and B notes), and whether the same element (A notes) is used across the two cry types. Secondly, observations of the cry variants under natural conditions were conducted to determine whether receivers are more likely to approach a caller producing SN compared to DN cries, and whether a group (including the caller) moves a larger distance in response to DN cries. Lastly, to verify behavioural observations pertaining to the function of the calls, and to prove that the same element is reused across two distinct contexts, we conducted playback experiments of natural, artificial, and control cries. We predicted receivers to approach the sound source when playing back natural, unmodified SN cries, and the group to fly off (following the putative signaller) in response to playbacks of natural DN cries. To verify that the A notes of SN and DN cries are acoustically equivalent and perceived as such, we played back artificial SN cries created out of DN cries (by deleting the cry's B notes; Fig. 1c). We predicted the same response to artificial SN cries as when exposed to natural SN cry playbacks. Lastly, in order to rule out that merely the cries' rhythmic patterns might account for their context specificity (i.e. any single or double-note motifs irrespective of the note types), control SN and DN cries were played back (Fig. 1d, e). Control SN cries were composed of B notes, and in control DN cries, each A note was substituted by a babbler contact call. If rhythmic patterns were solely responsible, we predicted equivalent responses to the control sequences as to the natural SN and DN variants.

Methods

Study site and species

We studied a free-living population of southern pied babblers at the Kuruman River Reserve (26°58'S, 21°49'E), located in the semi-arid Kalahari Desert of South Africa (for details on the habitat see [30, 39]). The research project was founded by ARR in 2003, and extensive life-history data are available for the individuals of the study population. All birds are habituated to human observers, enabling close observations up to 1m [26]. Unique combinations of three coloured and one numbered metal ring allow the identification of all babblers at the study site [26]. Pied babblers are sexually monomorphic, and individuals of the study population are sexed using DNA tests [25].

Acoustic analyses

Acoustic analyses were conducted to verify that the same acoustic element (A note) is used in SN and DN cries. Vocalisations were recorded using a Rode NTG-2 directional microphone coupled with a Rode Blimp Suspension Windshield (Rode microphones) and a Roland R-26 portable recorder (Roland Corporation) (sampling frequency 48kHz, 24-bit accuracy). Due to their high vocal activity and central role in leading the group to new areas, in addition to controlling for potential rank or age effects, only vocalisations of a group's dominant male were considered. In order to prevent pseudoreplication (and thus the generation of erroneous

P-value estimations), we only analysed one A note of a SN cry, as well as one A and one B note of a DN cry per individual [40]. For each of 10 dominant males high signal-to-noise-ratio elements were chosen randomly, and the following acoustic parameters were extracted using a custom-built script (E. Mandel-Briefer personal communication) in Praat (v. 5.1.03): element duration, bandwidth of fundamental frequency, maximum fundamental frequency, time-point of maximum fundamental frequency, time-point of maximum amplitude, as well as the frequency values at the upper limit of the first, second and third quartiles of energy. Additionally, we manually assessed the proportion of the element that exhibited clear voiced/periodic structures [41]. To assess the classification probabilities of element to note class (SN A note, DN A note, and DN B note), a cross-validated Discriminant Function Analysis (DFA) was conducted using the MASS package in R (v. 3.2.3) [42, 43]. With 3 element categories to be classified, the DFA created 2 discriminant functions by determining the linear combinations of the predictor variables that most adequately discriminate between the 3 categories [44]. Multivariate normality was assessed graphically, and only parameters with a variance inflation factor smaller than 5 were included in analyses, resulting in the exclusion of the frequency measurements at the first and third energy quartiles [45, 46]. Binomial tests were conducted to assess the significance of the DFA classification results (R stats package). According to the 3 element categories to be discriminated, a 33.3%-chance level of correct classifications was set.

Natural observations

Data collection

To investigate the context-specific production of, and the subsequent receiver response to the two types of recruitment cries, natural observations were conducted from January to April/May in 2014 and 2015. In 2014 we collected data on 19 babbler groups (average group size 6.2±2.3), and in 2015 on 18 groups (average group size 5.1±1.4). A group was followed in the evening for approximately 2h until it had settled in a night roost. The subsequent morning, the group was re-joined at the roost before dawn and followed for approximately 4h. Cry events of all male individuals, regardless of status, were recorded, documenting the identity of the caller, the cry type produced (SN or DN), and the distance moved by the caller. Additionally, we recorded the response of receivers, specifically whether the group approached the caller (yes/no), and the distance moved by the group in response to the cry. Movement was classified as receivers interrupting their foraging or resting activity and flying or fast moving on the ground to a new foraging or resting site. The moved distances by the caller and its group were measured using a global positioning data-logger (GPS-logger, eTrex 10, Garmin). We recorded the caller's location and the location of the density-based centre (barycentre) of the remaining group members during the cry production, as well as their locations after the group had switched to a new site in response to the cry. New sites were determined as locations where at least half of the group joined to rest or continued to forage. For SN cries, the location of a caller prior to and after the calling was generally identical. If no movement occurred within 5min after the cry production a distance of 0m was assigned.

Probability to approach the caller in response to SN and DN cries

To investigate whether receivers were more likely to approach a caller producing SN compared to DN cries, we fitted a binomial generalized linear mixed model (GLMM) with a logit-link function, with group approach specified as the response variable (0=no approach, 1=approach to caller). Caller identity nested within group identity was fitted as a random term, and cry pattern was the explanatory variable. The back-transformed model intercepts and the bootstrapped 95% confidence intervals (CI) were calculated to assess the percentage of occasions resulting in an approach to the caller in response to SN and DN cries respectively.

Distance moved by caller and receivers in response to SN and DN cries

To test whether the distance a caller and its receivers' moved, after producing and hearing SN and DN cries differed, we fitted a linear mixed model (LMM). The distance moved to a new foraging or resting site was fitted as the response variable, and square-root transformed to fulfil model assumptions. We fitted the cry pattern (SN/DN), identity of mover (caller/receiver) and the interaction between them as explanatory variables. The calling event, nested within caller identity and the caller's group was fitted as a random term. This allowed us to control for the dependency between the distance a caller and its receivers moved in response to one particular calling event (i.e. cry event 1 of individual X from group Y resulted in the caller moving *s* metres and its group moving *t* metres).

General statistical procedure

LMMs and GLMMs were fitted in R using the package lme4 [47]. Normality of the data was assessed graphically. Collinearity among predictor variables could be excluded in all models, since variance inflation factors were always substantially smaller than 5 [45]. Variance inflation factors were calculated using the vif.mer function designed for mixed effects models in R [48]. Potential overdispersion of the binomial GLMM was estimated by assessing the ratio of the sum of squared Pearson residuals to the degrees of freedom (each variance parameter representing one df), and could be ruled out [49]. Best models were chosen using the dredge function within the MuMIn package [50]. Based on the full model, the function selects the best model according to the Akaike's information criterion corrected for small sample sizes (AICc) [51]. A threshold difference (ΔAICc) of at least two to the next best model was set. Significances for relevant fixed effects were calculated based on bootstrapping methods applying 10,000 iterations using the pbkrtest package, which reports the test statistic (PB value) and the corresponding P-value [52]. To determine differences between the levels of significant factors, we assessed whether the bootstrapped CIs of given contrasts intersected with zero, with CIs not crossing zero representing significant contrasts. Back-transformed model estimates were plotted using ggplot2 [53] and gridExtra [54].

Playback experiments

Playback experiments on 10 babbler groups were conducted between March and May 2015. For consistency and due to their high vocal activity, we only used vocalisation of a group's dominant male to create the playback stimuli. High signal-to-noise-ratio vocalisations were selected and edited using Audition CS6 (Adobe, sampling frequency 48kHz, 24-bits accuracy). To verify the recruitment cries' context specificity, in each of 10 test groups, a natural SN cry and a natural DN cry were played back (Fig. 1a, b). To investigate whether the SN and DN cry are composed of the same acoustic element (the A note), an artificial SN cry was created by deleting each B note of the natural DN cry (Fig. 1c). The inter-element distance between the remaining A notes was adjusted to that found between A notes of the natural SN cry. To rule out that any single or double-note motifs, irrelevant of the note type, might elicit a response as observed to natural SN and DN cries, control stimuli were created. A control SN cry was created by deleting each A note within the natural DN cry, with the inter-element distance between the remaining B notes again matched to the distance found between the A notes of the natural SN cry (Fig. 1d). A control DN cry was created by substituting each natural DN cry's A note with a chuck vocalisation (babbler foraging contact call [26]) of the same individual, keeping the same distance between the replaced element and the B notes (Fig. 1e). Since chuck vocalisations can exhibit substantial acoustic variation that correlates with the type and quality of the food source a babbler is processing [29], we repeated the same chuck exemplar so as to avoid presenting variable and contextually inconsistent chucks within one cry.

Thus, 5 stimuli were created from one SN cry, one DN cry, and one chuck vocalisation, each originating from the same dominant individual in each of the 10 test groups. In cases where elements were replaced, the amplitude of the new element was matched to the one of the substituted element. All 5 stimuli cries for one group were adjusted to have approximately the same duration (4.7±0.5sec), with some cries being shortened and others being extended by duplicating elements within a cry. All stimuli were normalised and played back at a naturally occurring amplitude (~73dB at 4m distance, measured using a Cirrus CR261 sound level meter). Each cry was broadcast once using a portable speaker (Anchor AN-30) coupled with an iPod 3 (Apple Inc.). Each of the 10 groups received all 5 stimuli in a randomised order. A maximum of 2 stimuli were played back in one morning, with at least 2 days break in-between subsequent playback sessions to minimise the risk of habituation to playbacks. Since vocalisations of a group's dominant male were broadcast, and recruitment cries are individually distinct [38], stimuli were played back from the location of the dominant male at the time of the playback. This served to prevent reactions to the stimuli based solely on the incongruence between the location of the playback and the location of the dominant male. Accordingly, the loudspeaker was placed below or next to a group's dominant male, with the rest of the group foraging in at least 20m distance to the male (mean 25±4.5m). We recorded the proportion of the group approaching within 5m to the loudspeaker, as well as the distance the whole group moved 5 and 10min after the playback using a GPS data-logger. To investigate differences in the groups' responses to the playback stimuli, two-tailed, paired Friedman and post-hoc Wilcoxon tests, with asymptotic P-values adjusted using the "holm"-method, were carried out in R using the MASS package [42].

Results

Acoustic analyses

In accordance with our predictions, a cross-validated DFA ($N_{elements}$=30, $N_{individuals}$=10) indicated that B notes were structurally distinct from A notes, never being misclassified as A notes and vice versa (correct classification: B notes=100%, P<0.001; Fig. 2). A notes originating from SN and DN cries were indistinguishable from each other (correct classifications: SN A note=40%, DN A note=30%, all P>0.5; with the remaining percentage of elements being classified as A note of the opposing cry type).

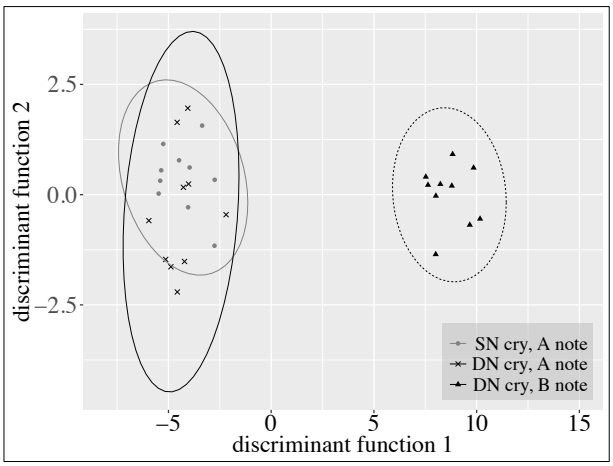


Figure 2. DFA output on SN cry A notes, and DN cry A and B notes of 10 dominant males. Circles represent clusters assuming a multivariate normal distribution with a confidence level of 95%.

Natural observations

Probability of approach in response to SN and DN cries

Receivers were more likely to approach callers producing SN cries than DN cries (PBtest=44.5, P<0.001, N=152 cries of 31 individuals in 20 groups). For SN cries, receivers approached the caller in 63.7% (CIs: 53.7-74.2%) of occasions. In contrast, receivers approached the caller in response to DN cries in only 12.5% (CIs: 5.6-19.6%) of occasions.

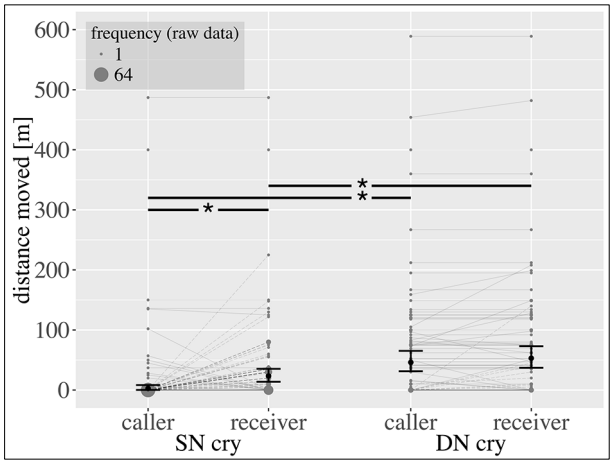


Figure 3. Moved distance of callers and receivers in response to SN and DN cries. Bars illustrate the 95% confidence intervals, and points show the medians of the back-transformed data. Pale dots show the raw data, with the size corresponding to the frequency of occurrences. Lines connect calling events, i.e. the distance a caller and its group moved in a specific calling event. N=295 observations originating from 153 calling events of 31 individuals in 20 groups. Asterisks indicate significant contrasts according to the CIs. Although not part of the statistical model investigating differences in the moved distance, but in order to provide a more comprehensive representation of the data, occasions were receivers approached the caller are additionally illustrated as dashed lines.

Distance moved in response to SN and DN cries

How far callers and their receivers moved after producing and hearing recruitment cries differed based on the cry type produced (cry pattern [SN *vs*. DN] x subject [caller *vs*. receiver]: PBtest=16.7; cry pattern: PBtest=39.4; subject: PBtest=44.2, all P<0.001, N=295 observations originating from

153 calling events of 31 individuals in 20 groups). Specifically, when producing SN cries, callers moved less than receivers and generally remained stationary (64 out of 79 occasions), which is in accordance with our results demonstrating receivers approaching the caller in response to SN cries. When producing DN cries, callers and receivers moved the same distance. Furthermore, receivers (and callers) moved a greater distance in response to DN compared to SN cries (Fig. 3).

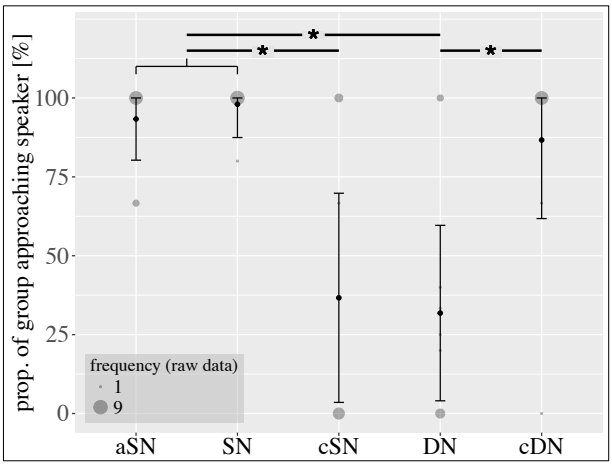


Figure 4. Proportion of the group approaching the loudspeaker in response to natural SN cries (SN), natural DN cries (DN), control SN cries (cSN), control DN cries (cDN), and artificial SN cries (aSN). Bars illustrate the 95% confidence intervals, and points show the medians over 10 groups. Pale dots show the raw data, with the size corresponding to the frequency of occurrences. Asterisks indicate significant differences.

Playback experiments

The proportion of group members approaching the loudspeaker differed in response to playbacks of natural, artificial and control SN and DN cries (χ^2_4=19.8, P<0.001, N=10 groups; Fig. 4). In line with our predictions, artificial SN cries elicited the same response as natural SN cries (V=8, P=0.41, N=10). Fewer individuals approached the loudspeaker in response to natural DN cries compared to playbacks of natural SN (V=36, P=0.01, N=10) or artificial SN cries (V=43, P=0.02, N=10). Ruling out rhythmicity effects, neither control SN cries or control DN cries elicited the same response as their natural (or artificial) counterparts (control SN *vs*. natural SN: V=28; control SN *vs*. artificial SN: V=28; control DN *vs*. natural DN: V=1; all P=0.02, N=10). No differences in responses to the playback stimuli were found regarding the distance a group moved in 5 or 10 minutes after playback onset (5min.: χ^2_4=2.2, P=0.69; 10min.: χ^2_4=5.17, P=0.27; N=10 groups).

Discussion

Here, we provide empirical evidence for combinatoriality in the internal structure of pied babbler recruitment cries. Acoustic analyses and playback experiments confirm babblers reuse the same acoustic element across the single-note motifs [$(A)^n$] of SN cries and the double-note motifs [$(AB)^n$] of DN cries. The A notes of both cries were statistically indistinguishable but distinct from B notes, and receivers engaged in the same heightened approach response to playbacks of natural and artificial SN cries. Neither playbacks of control SN or control DN cries elicited the same response as natural SN or natural DN cries, ruling out the alternative explanation that the contextual information of the two recruitment cries might simply be encoded in the cries' rhythmic motif pattern (i.e. random single- or double-note motifs).

Regarding the function of SN and DN cries, our observational data and behavioural responses to playbacks verify the prediction that SN cries function to recruit group members to the caller's current location. The observational data also confirm our hypothesis that DN cries initiate a cohesive movement of the group, but interestingly, playbacks of DN cries did not fully affirm this, as they failed to elicit a greater movement in comparison to SN cries. Instead we found no statistical difference between the playback treatments regarding the distance travelled, indicating that, while receivers travelled toward the sound source in response to playbacks of both natural and artificial SN cries, for the other treatments there was no directional preference in movement. The absence of an increased movement of receivers in response to playback of DN cries might suggest DN cries do not function to induce cohesive movement over longer distances. However, in light of the observational data, we suggest this is more likely a consequence of our experimental setup. All playback stimuli were broadcast from the dominant male's current location. While simulating his likely motivation to move by playing back DN cries, it was not feasible to also visually simulate the male's subsequent departure, which necessarily accompanies the production of DN cries. Accordingly, the lack of a movement response by receivers is likely a consequence of the simulated caller, although allegedly announcing a movement, then not moving, and receivers missing key visual information pertaining to the relevant direction.

These methodological issues notwithstanding, our work suggests that the super-structure of both cry types (introduced by a wind-up, repetitions of reoccurring motifs, similar average duration and loudness) likely conveys the same intention of the signaller to recruit group members, with the internal motif pattern determining the precise form of recruitment. A/single-note motifs appear to address receivers to approach the caller, whereas AB/double-note motifs intensify recruitment from requesting an approach to following the signaller over longer distances. Accordingly, the B note might represent an acoustic modifier, which alone is devoid of function, but when combined with an A note alters/intensifies the A note's meaning. From a signaller and receiver perspective these two forms of recruitment are likely more efficient than having to reuse the same signal (i.e. SN cries) many times to recruit group members over longer distances. Firstly, signallers reduce the risk of attracting predators through avoiding repeated, loud vocal announcements. Secondly, given babblers are weak flyers and short flights bear high energetic costs [55], possessing a specific vocalisation that signals the likely motivation to move far permits a smoother, more flexible follow of the caller instead of repeated, costly approaches.

Although further experimental work is necessary to clarify how babblers process the information in the combinatorial structure of these two recruitment cries, our work provides further support for the implementation of combinatorial mechanisms as a way to likely increase communicative output in non-human communication systems. Parallels can be found in primate alarm call systems such as in Campbell's monkeys, where an acoustic modifier appears to broaden the meaning of predator-specific alarm calls [21, 22]. In contrast with Campbell's monkeys, however, pied babblers apply the B-modifier in only one instance (i.e. DN cries), raising the question why, instead of A- and AB-motifs, babblers do not simply apply stand-alone A- and B-motifs. We propose that even though A and B notes are acoustically distinguishable, these differences are subtle and potentially blur when broadcast

over long distances, increasing the risk of signal misperception. Stringing sounds together, on the other hand, can act to create more easily distinguishable signals [7], and counter the possibility of signal degradation hampering discriminability between otherwise similar sounds (i.e. A and B notes).

To conclude, pied babbler recruitment cries represent a novel type of combinatorial structuring, with embedded acoustic variation specifying the signal's function, and add further insight into the diversity and distribution of combinatorial mechanisms outside of humans [10-13]. Our work supports the dominant theory that meaningful vocal combinations emerged to overcome production and perceptional limitations experienced by non-open ended learners, ultimately enhancing signal transmission and increasing a species' communicative output [7, 18]. Exactly what combinatorial mechanisms an animal implements (phoneme-like structuring, intensifications, affixation, compositionality etc. [14, 18, 21, 23, 24]), among other factors, may well depend on the context (social or predatory) that vocalisations are given in and the potential costs implemented by misperception, as well as whether signals are bound to short-range or long-range communication [14]. Accordingly, further comparative work is essential in order to more accurately identify the ecological and social conditions that likely drive the emergence and, particularly, the variation of combinatorial mechanisms in non-human animals.

Ethics statement. The study was conducted under the permission of the ethical committee for animal research of the University of Cape Town and the Northern Cape Conservation Authority, South Africa.

Acknowledgements. We thank T. Clutton-Brock, the Kalahari Research Trust, the families de Bruin and Kotze for access to land and logistics; all past and present researchers and assistants at the Pied Babbler Research Project for maintaining habituation of the study population and collecting life history data.

Author contribution. SE, ARR, and SWT designed the research; SE performed the research and analysed the data; SE and SWT wrote the paper; and ARR and MM contributed in discussions and revisions. The authors declare no competing interests.

Funding statement. Funding was provided by the Forschungskredit of the University of Zurich Grants 57191601 (to SE) and FK-14-077 (to SE), Swiss National Science Foundation Grants P1ZHP3_151648 (to SE) and 31003A_153065 (to SWT), and the Claraz Stiftung (to SWT).

References

[1] Hockett CF. 1960 The Origin of Speech. *Sci. Am.* **203**, 88-111.
[2] Chomsky N. 1981 Knowledge of Language: Its Elements and Origins. *Philos. T. R. Soc. B* **295**, 223-234. (doi:10.1098/rstb.1981.0135)
[3] Humboldt Wv. 1999 *On Language. Translated by Peter Heath*. Cambridge: Cambridge University Press.
[4] Hauser MD, Chomsky N, Fitch WT. 2002 The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* **298**, 1569-1579. (doi:10.1126/science.298.5598.1569)

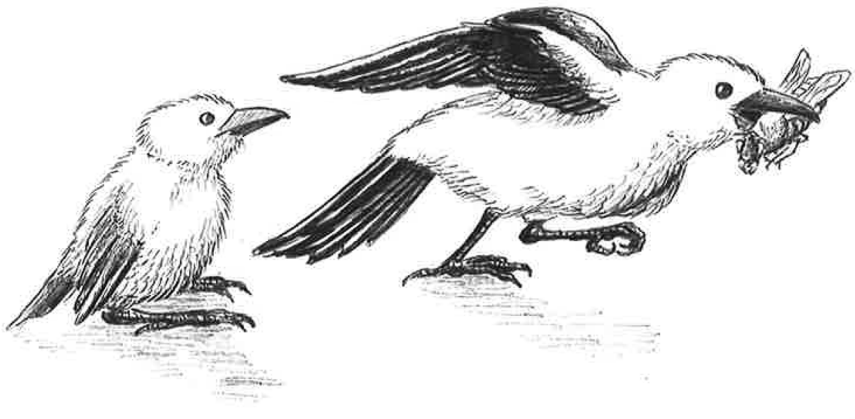
[5] Christiansen MH, Kirby S. 2003 Language evolution: consensus and controversies. *Trends Cogn. Sci.* **7**, 300-307. (doi:10.1016/s1364-6613(03)00136-0)
[6] Nowak MA, Krakauer DC. 1999 The evolution of language. *Proc. Natl. Acad. Sci. USA* **96**, 8028-8033. (doi:10.1073/pnas.96.14.8028)
[7] Nowak MA, Krakauer DC, Dress A. 1999 An error limit for the evolution of language. *Proc. R. Soc. B* **266**, 2131-2136. (doi:10.1098/rspb.1999.0898)
[8] Nowak MA, Komarova NL. 2001 Towards an evolutionary theory of language. *Trends Cogn. Sci.* **5**, 288-295. (doi:10.1016/S1364-6613(00)01683-1)
[9] Rendall D. 2013 Q&A: Cognitive ethology - inside the minds of other species. *BMC Biol.* **11**, 1-6. (doi:10.1186/1741-7007-11-108)
[10] Marler P. 1977 *The structure of animal communication sounds. Recognition of complex acoustic signals: report of Dahlem workshop*. Berlin: Abakon Verlagsgesellschaft.
[11] Hurford J. 2011 Linguistics from an Evolutionary Point of View. In *Handbook of the Philosophy of Science: Linguistics* (eds. R Kempson, M Fernando & N Ashe), pp. 473-498. Amsterdam: Elsevier Science.
[12] Collier K, Bickel B, van Schaik CP, Manser MB, Townsend SW. 2014 Language evolution: syntax before phonology? *Proc. R. Soc. B* **281**, 20140263. (doi:10.1098/rspb.2014.0263)
[13] Zuberbühler K. 2015 Linguistic capacity of non-human animals. *WIREs Cogn. Sci.* **6**, 313-321. (doi:10.1002/wcs.1338)
[14] Engesser S, Crane JM, Savage JL, Russell AF, Townsend SW. 2015 Experimental Evidence for Phonemic Contrasts in a Nonhuman Vocal System. *PLoS Biol.* **13**, e1002171. (doi:10.1371/journal.pbio.1002171)
[15] Candiotti A, Zuberbühler K, Lemasson A. 2012 Context-related call combinations in female Diana monkeys. *Anim. Cogn.* **15**, 327-339. (doi:10.1007/s10071-011-0456-8)
[16] Jansen DA, Cant MA, Manser MB. 2013 Segmental concatenation of individual signatures and context cues in banded mongoose (*Mungos mungo*) close calls. *BMC Biol.* **10**, 97. (doi:10.1186/1741-7007-10-97)
[17] Coye C, Zuberbühler K, Lemasson A. 2016 Morphologically structured vocalizations in female Diana monkeys. *Anim. Behav.* **115**, 97-105. (doi:10.1016/j.anbehav.2016.03.010)
[18] Arnold K, Zuberbühler K. 2006 Language evolution: Semantic combinations in primate calls. *Nature* **441**, 303. (doi:10.1038/441303a)
[19] Arnold K, Zuberbühler K. 2008 Meaningful call combinations in a non-human primate. *Curr. Biol.* **18**, R202-R203. (doi:10.1016/j.cub.2008.01.040)
[20] Arnold K, Zuberbühler K. 2012 Call combinations in monkeys: Compositional or idiomatic expressions? *Brain Lang.* **120**, 303-309. (doi:10.1016/j.bandl.2011.10.001)
[21] Ouattara K, Lemasson A, Zuberbühler K. 2009 Campbell's Monkeys Use Affixation to Alter Call Meaning. *PLoS ONE* **4**, e7808. (doi:10.1371/journal.pone.0007808)
[22] Coye C, Ouattara K, Zuberbühler K, Lemasson A. 2015 Suffixation influences receivers' behaviour in non-human primates. *Proc. R. Soc. B* **282**, 20150265. (doi:10.1098/rspb.2015.0265)
[23] Engesser S, Ridley AR, Townsend SW. 2016 Meaningful call combinations and compositional processing in the southern pied babbler. *Proc. Natl. Acad. Sci. USA* **113**, 5976-5981. (doi:10.1073/pnas.1600970113)
[24] Suzuki TN, Wheatcroft D, Griesser M. 2016 Experimental evidence for compositional syntax in bird calls. *Nat. Commun.* **7**, 10986. (doi:10.1038/ncomms10986)
[25] Radford AN, Ridley AR. 2006 Recruitment Calling: A Novel Form of Extended Parental Care in an Altricial Species. *Curr. Biol.* **16**, 1700-1704. (doi:10.1016/j.cub.2006.06.053)
[26] Radford AN, Ridley AR. 2008 Close calling regulates spacing between foraging competitors in the group-living pied babbler. *Anim. Behav.* **75**, 519-527. (doi:10.1016/j.anbehav.2007.05.016)
[27] Raihani NJ. 2008 Cooperation and conflict in pied babblers, University of Cambridge, PhD.
[28] Bell MBV, Radford AN, Rose R, Wade HM, Ridley AR. 2009 The value of constant surveillance in a risky environment. *Proc. R. Soc. B* **276**, 2997-3005. (doi:10.1098/rspb.2009.0276)
[29] Golabek KA. 2010 Vocal communication and the facilitation of social behaviour in the southern pied babbler (*Turdoides bicolor*), University of Bristol, PhD.
[30] Hollén LI, Bell MBV, Russell A, Niven F, Ridley AR, Radford AN. 2011 Calling by Concluding Sentinels: Coordinating Cooperation or Revealing Risk? *PLoS ONE* **6**, e25010. (doi:0.1371/journal.pone.0025010)
[31] Thompson A. 2012 Post-fledging begging and development in Southern Pied Babblers (*Turdoides bicolor*), University of Cape Town, PhD.
[32] Golabek KA, Radford AN. 2013 Chorus-call classification in the southern pied babbler: multiple call types given in overlapping contexts. *Behaviour* **150**, 691-712. (doi:10.1163/1568539X-00003074)
[33] Ridley AR, Raihani NJ. 2007 Variable postfledging care in a cooperative bird: causes and consequences. *Behav. Ecol.* **18**, 994-1000. (doi:10.1093/beheco/arm074)

[34] Nelson-Flower MJ, Hockey PAR, O'Ryan C, Raihani NJ, du Plessis MA. 2011 Monogamous dominant pairs monopolize reproduction in the cooperatively breeding pied babbler. *Behav. Ecol.* **22**, 559-565. (doi:10.1093/beheco/arr018)
[35] Humphries D. 2012 The mechanisms and function of social recognition in the cooperatively breeding Southern pied babbler, *Turdoides bicolor*, Macquarie University, PhD.
[36] Ridley AR, Raihani NJ. 2006 Facultative response to a kleptoparasite by the cooperatively breeding pied babbler. *Behav. Ecol.* **18**, 324-330. (doi:10.1093/beheco/arl092)
[37] Golabek KA, Ridley AR, Radford AN. 2012 Food availability affects strength of seasonal territorial behaviour in a cooperatively breeding bird. *Anim. Behav.* **83**, 613-619. (doi:10.1016/j.anbehav.2011.11.034)
[38] Humphries D, Finch FM, Bell MB, Ridley AR. 2016 Vocal Cues to Identity: Pied Babblers Produce Individually Distinct But Not Stable Loud Calls. *Ethology* **122**, 609--619. (doi:10.1111/eth.12 508)
[39] Raihani NJ, Ridley AR. 2007 Adult vocalizations during provisioning: offspring response and postfledging benefits in wild pied babblers. *Anim. Behav.* **74**, 1303-1309. (doi:10.1016/j.anbehav.2007.02.025)
[40] Sharp SP, McGowan A, Wood MJ, Hatchwell BJ. 2005 Learned kin recognition cues in a social bird. *Nature* **434**, 1127-1130. (doi:10.1038/nature03522)
[41] Fitch WT, Neubauer J, Herzel H. 2002 Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production. *Anim. Behav.* **63**, 407-418. (doi:10.1006/anbe.2001.1912)
[42] Venables WN, Ripley BD. 2002 *Modern Applied Statistics with S*. New York: Springer.
[43] R-Core-Team. 2014 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
[44] Mundry R, Sommer C. 2007 Discriminant function analysis with nonindependent data: consequences and an alternative. *Anim. Behav.* **74**, 965-976. (doi:10.1016/j.anbehav.2006.12.028..)
[45] Zuur A, Ieno EN, Walker N, Saveliev AA, Smith GM. 2009 *Mixed Effects Models and Extensions in Ecology with R*. New York: Springer.
[46] Fox J, Weisberg W. 2011 *An {R} Companion to Applied Regression, Second Edition*. Thousand Oaks CA: Sage.
[47] Bates D, Maechler M, Bolker B, Walker S. 2014 Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1-48.
[48] Frank AF. 2011 R-hacks/mer-utils.R. https://github.com/aufrank/R-hacks/blob/master/mer-utils.R. Accessed 29 Apr 2016.
[49] Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MH, White JS. 2009 Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.* **24**, 127-135. (doi:10.1016/j.tree.2008.10.008)
[50] Barto'n K. 2014 MuMIn: multi-model inference. R package version 1.10.5. http://CRAN.R-project.org/package=MuMIn. Accessed 29 Apr 2016.
[51] Burnham KP, Anderson DR. 2002 *Model selection and multimodel inference: a practical information-theoretic approach*. New York: Springer.
[52] Halekoh U, Højsgaard S. 2014 A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models - TheRPackagepbkrtest. *J. Stat. Softw.* **59**, 1-32. (doi:10.18637/jss.v059.i09)
[53] Wickham H. 2009 *ggplot2: elegant graphics for data analysis*. New York: Springer.
[54] Auguie B. 2015 gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.0.0. http://CRAN.R-project.org/package=gridExtra. Accessed 29 Apr 2016.
[55] Nudds RL, Bryant DM. 2000 The energetic cost of short flights in birds. *J. Exp. Biol.* **203**, 1561-1572.

CHAPTER 5

Element repetition rates encode functionally distinct information in pied babbler 'clucks' and 'purrs'

To be submitted



Element repetition rates encode functionally distinct information in pied babbler 'clucks' and 'purrs'

Sabrina Engesser [a], Amanda R. Ridley [b,c], Simon W. Townsend [a,d]

[a] Animal Behaviour, Department of Evolutionary Biology and Environmental Studies, University of Zurich, 8057 Zurich, Switzerland

[b] Centre for Evolutionary Biology, School of Animal Biology, The University of Western Australia, Crawley WA 6009, Australia

[c] Percy FitzPatrick Institute, University of Cape Town, Rondebosch 7701, South Africa

[d] Department of Psychology, University of Warwick, Coventry CV4 7AL, United Kingdom

Abstract

Human language is a recombinant system that achieves its productivity through the combination of a limited set of sounds. Research investigating the evolutionary origin of this generative capacity has generally focused on the capacity of non-human animals to combine different types of discrete sounds to encode new meaning, with less emphasis on meaning-differentiating mechanisms achieved through potentially more simple temporal modifications within a sequence of repeated sounds. Here we show that pied babblers (*Turdoides bicolor*) generate two functionally distinct vocalisations composed of the same sound type, which can only be distinguished by the number of repeated elements. Specifically, babblers produce extended 'purrs', composed of, on average, around 14 element repetitions when drawing young offspring to a food source, and truncated 'clucks' composed of a fixed number of 2-3 elements when collectively mediating imminent changes in foraging site. We propose that meaning-differentiating temporal structuring might be a much more widespread combinatorial mechanism than currently recognised, and is likely of particular value for species with limited vocal repertoires in order to increase their communicative output.

Keywords: element repetition - temporal structure - animal communication - combinatoriality - language evolution - *Turdoides bicolor*

Introduction

Animals employ various mechanisms to communicate and transfer information [1]. Combinatorial operations, whereby animals combine discrete sounds together into larger meaningful structures, have received particular recent interest in part due to their similarities to layers of articulation in language [2]. Animals further encode information through modifying the temporal arrangement of repeated elements within a sequence [3]. Such temporal structuring has been shown to convey information on motivation or threat-levels experienced by an individual during aggressive or predatory encounters [3-6] (though see also [7]). In these instances, temporal modifications generally constitute a graded system correlating with a signaller's arousal level, and typically take the form of changes in the number/rate of repeated elements or in inter-element intervals [3-6]. However, work on the alarm calls of colobus monkeys (*Colobus guereza & polykomos*) and the social vocalisations of Mexican free-tailed bats (*Tadarida brasiliensis*) have shown temporal changes can also encode both distinct predator-specific and behaviour-specific information [8-10]. These results suggest that, besides transferring graded information, temporal modifications can encode more discrete, categorical information; a feature more commonly associated with the combination of different sound types [11].

Here, we investigate whether similar temporally induced meaning-differentiation also exists outside of mammals, in the discrete vocal system of a social bird, the pied babbler (*Turdoides bicolor*). Babblers live in stable, territorial groups of up to 15 individuals [12]. They are weak flyers and spend most of their time on the ground searching for prey [12]. Group members stay close together throughout the day, and move cohesively between foraging sites, which are unevenly distributed throughout their territory [12]. Longer movements are frequently associated with 'recruitment cries' produced by a single individual (often the dominant male) to prompt the group to follow [13]. Group members also roost together in the same tree overnight [14].

Previous work demonstrated that during nestling provisioning and when attracting fledglings towards food patches, babblers produce soft 'purr' vocalisations, composed of an extended number of up to ~30 repetitions of the same acoustic element [12, 15, 16]. Preliminary work indicated that babblers also produce truncated variants of purrs with a

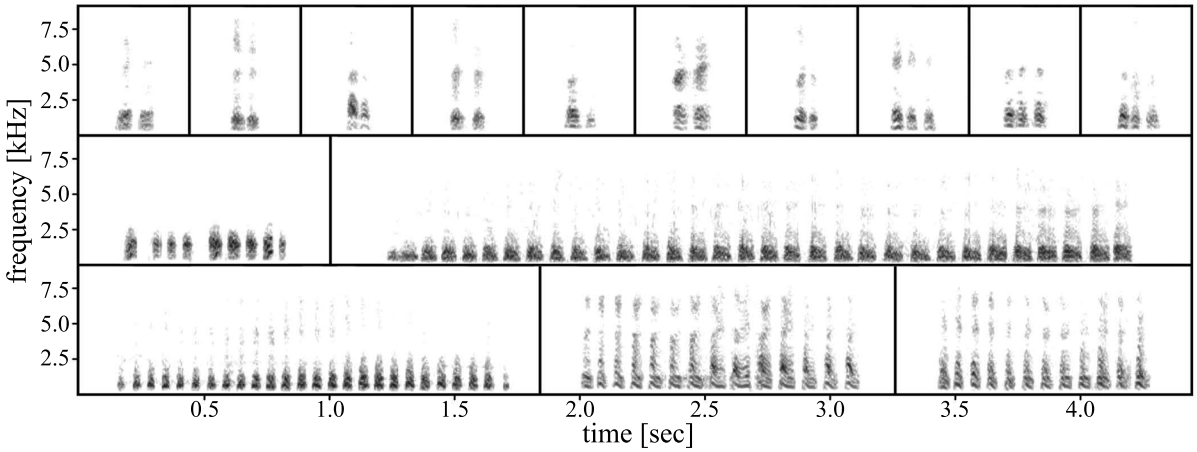


Figure 1. Spectrograms of clucks (top row) and purrs (lower 2 rows) of different individuals.

considerably reduced number of 2-3 elements, named here as 'clucks' (Fig. 1). Initial observations indicated that clucks are jointly produced by multiple group members (usually over several minutes) before groups move to the first foraging site in the morning, or generally before switching between foraging sites. Crucially, the number of clucking individuals seems to impact a group's decision to move between foraging sites, with the number of callers correlating with the distance a group eventually moves to a new site (SE personal observation). Accordingly, and in contrast to recruitment cries (being only produced by a single bird), clucks may represent a more egalitarian decision-making mechanism, linking individual preferences to group travel processes.

In order to investigate the temporal features of clucks and purrs, we examined, using acoustic analyses, whether the same element is used within and across the two call types, and whether their temporal distinctiveness is determined by the number of repeated elements. To verify the functional specificity of clucks (in comparison to purrs), behavioural observations were conducted in which the context of cluck production, and responses to clucks were recorded. Critically, we expected a group to move a longer distance when more group members were clucking prior to the movement.

Methods

Data were collected between January and May 2015 on 15 groups of free-living, habituated pied babblers at the Kuruman River Reserve, South Africa (26°58'S, 21°49'E) (mean group size=5.1±0.3, range 3-7). To verify the acoustic similarity among cluck and purr elements, in each of 11 groups we acoustically analysed a purr and a cluck vocalisation (see supplementary for detailed information), and recorded the number of element repetitions. Natural observations were conducted to investigate the effect of the number of clucking individuals on the distance a group subsequently moved, i) when leaving the roost at dawn, and ii) during foraging. Additionally, we documented a group's size and its total clucking duration to rule out a potential group size effect, and the possibility that longer calling durations per se might result in more calling individuals. A cross-validated Discriminant Function Analysis (cDFA) and a binomial test were conducted to investigate the acoustic similarity among cluck and purr elements. Linear mixed models were fitted and model selection based Akaike's information criterion corrected for small sample sizes (AICc) were applied to investigate the effect of the predictor variables on the distance a group moved after clucking. See supplementary material for more details on methods.

Results

The elements comprising clucks and purrs could neither be discriminated within or across the two calls (cDFA: correct classifications=25%, P=1, $N_{elements}$=44, N_{groups}=11; Fig. 2a). Clucks were composed of 2-3 repeated elements (mean=2.5±0.5), and purrs of 5-32 (mean=13.6±8.3), and could clearly be distinguished by the number of elements (paired, two-tailed Wilcoxon signed rank test: V=66, P<0.01). Behavioural observations revealed groups travelled a greater distance to the first foraging site in the morning, and between foraging

sites during the day, respectively, when more individuals were clucking prior to movement (roosting: PBtest=21.6, N=36; foraging: PBtest=15.0, N=98; all P<0.001; Fig. 2b & Table S1). Group size and the overall clucking duration of a group did not affect subsequent movements, as they were not retained during model selection (see supplementary material).

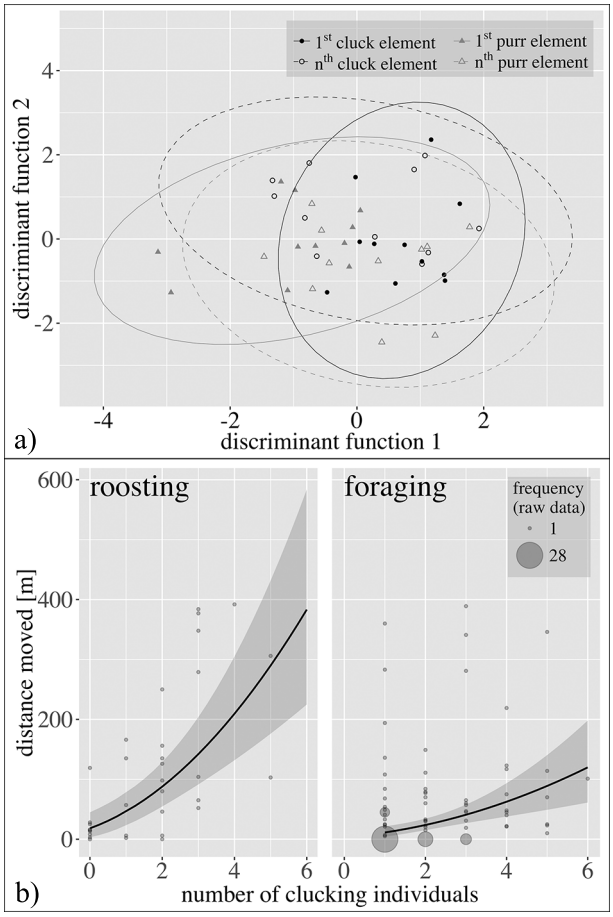


Figure 2. a) DFA output on cluck and purr elements. Circles represent clusters assuming a multivariate normal distribution with a confidence level of 95%. b) Distance a group moved to a foraging site as a function of the number of clucking individuals, in the morning when leaving the roost (left panel), and during foraging (right panel). Shaded areas illustrate the 95% confidence intervals. Dots show the raw data with the size corresponding to the frequency of occurrences.

Discussion

Acoustic analyses confirmed that babbler clucks and purrs are composed of the same acoustic elements, with only the number of element repetitions distinguishing the two calls. Previous work has demonstrated that purrs serve to attract dependent offspring towards a food source [12, 15, 16]. Here, we show that clucks and purrs, although composed of identical elements, are produced in different contexts. When produced at the roost tree in the morning, babblers flew further to a foraging site when more group members were clucking. In contrast, when no clucks were emitted, groups generally started to forage in the immediate vicinity of the roost tree. This result was supported by observations conducted during foraging, showing when more individuals were clucking prior to switching foraging sites, the distance moved between the two sites was larger. Our work therefore suggests that clucks might function to mediate imminent movement. However, the factors that drive their production need further investigation. Clucks are potentially produced to request movement when individual foraging patches are depleted, or when the location of the roost tree is sub-optimal as a suitable foraging site. A similar mechanism is implemented by the socially foraging meerkat (*Suricata suricatta*), where 'move' calls produced by multiple individuals may indicate a depletion of the foraging site, and induce group movement [17]. In line with this, babbler clucks might function as a vocal tool to gather information on individual preferences or assessments of the group's current location [17], with movement decisions being modulated through attending to the number of calling individuals. Positive feedback mechanisms might explain the increase in the distance moved with a greater number of clucking individuals [18]. Specifically, the more individuals exhibiting a particular

preference, the higher the likelihood that further individuals join in [18], or alternatively, the greater the magnitude of the particular behaviour. However, the intensity of conflicting interests might also be context specific, diminishing or changing under certain circumstances [18]. When already foraging, babblers might have stronger tendencies to stay, or switch sites respectively, since individual foraging patches and needs might vary more during foraging in comparison to after roosting where all group members equally require food intake. This may explain the less pronounced relationship between clucking individuals and distance moved during foraging compared to after roosting.

Although further systematic observations and experiments are needed to verify the hypothesised function of clucks, specifically how the number of clucking individuals translates into the distance moved, we show here that element repetition rates generate two functionally distinct signals: the 'truncated' clucks and the 'extended' purrs. While recent work has begun to address non-human animal vocal combinatorial abilities, the main focus has been on the ability to combine different sound types to encode new, discrete information [1, 2, 11]. However, temporal mechanisms have largely been neglected, potentially because they do not present an obvious analogue to language's articulatory layers. In line with previous work in both primates and bats [8-10] we show, that rather than encoding quantitative information related to arousal levels, modifications of temporal patterns can additionally transfer qualitatively different information. While such temporal modifications seem relatively simple, we propose that they may still play an important role in increasing a species' vocal repertoire, and similarly to combinations of discrete sounds, temporal operations might be a widespread mechanism applied by various animals to encode diverse information sets.

Ethics statement. Permission was provided by the ethical committee for animal research of the University of Cape Town and the Northern Cape Conservation Authority, South Africa.

Acknowledgements. We thank the Kalahari Research Trust, T. Clutton-Brock, M. Manser, the de Bruins and Kotzes for logistics and access to land; researchers and assistants at the babbler project; E. Mandel-Briefer for help on acoustics; M. Manser for support; and M. Manser & C. Bousquet for comments on previous versions.

Author contribution. SE, ARR & SWT designed research; SE performed research and analysed data; SE & SWT wrote paper. The authors declare no competing interests.

Funding statement. Funding was provided by the Forschungskredit of the University of Zurich Grants 57191601 (to SE) and FK-14-077 (to SE), Swiss National Science Foundation Grants P1ZHP3_151648 (to SE) and 31003A_153065 (to SWT), and the Claraz Stiftung (to SWT).

References

[1] Marler P. 1977 *The structure of animal communication sounds. Recognition of complex acoustic signals: report of Dahlem workshop*. Berlin: Abakon Verlagsgesellschaft.

[2] Collier K, Bickel B, van Schaik CP, Manser MB, Townsend SW. 2014 Language evolution: syntax before phonology? *Proc. R. Soc. B* **281**, 20140263. (doi:10.1098/rspb.2014.0263)
[3] Kershenbaum A, Blumstein DT, Roch MA, Akcay C, Backus G, Bee MA, Bohn K, Cao Y, Carter G, Casar C, et al. 2014 Acoustic sequences in non-human animals: a tutorial review and prospectus. *Biol. Rev. Camb. Philos. Soc.* **91**, 13-52. (doi:10.1111/brv.12160)
[4] DuBois AL, Nowicki S, Searcy WA. 2009 Swamp sparrows modulate vocal performance in an aggressive context. *Biol. Lett.* **5**, 163-165. (doi:10.1098/rsbl.2008.0626)
[5] Rek P, Osiejuk TS. 2012 Temporal patterns of broadcast calls in the corncrake encode information arbitrarily. *Behav. Ecol.* **24**, 547-552. (doi:10.1093/beheco/ars196)
[6] Wheatcroft D. 2015 Repetition rate of calls used in multiple contexts communicates presence of predators to nestlings and adult birds. *Anim. Behav.* **103**, 35-44. (doi:10.1016/j.anbehav.2015.02.009)
[7] Antunes R, Schulz T, Gero S, Whitehead H, Gordon J, Rendell L. 2011 Individually distinctive acoustic features in sperm whale codas. *Anim. Behav.* **81**, 723-730. (doi:10.1016/j.anbehav.2010.12.019)
[8] Bohn KM, Schmidt-French B, Ma ST, Pollak GD. 2008 Syllable acoustics, temporal patterns, and call composition vary with behavioral context in Mexican free-tailed bats. *J. Acoust. Soc. Am.* **124**, 1838-1848. (doi:10.1121/1.2953314)
[9] Schel AM, Tranquilli S, Zuberbühler K. 2009 The alarm call system of two species of black-and-white colobus monkeys (*Colobus polykomos* and *Colobus guereza*). *J. Comp. Psychol.* **123**, 136-150. (doi:10.1037/a0014280)
[10] Schel AM, Candiotti A, Zuberbühler K. 2010 Predator-deterring alarm call sequences in Guereza colobus monkeys are meaningful to conspecifics. *Anim. Behav.* **80**, 799-808. (doi:10.1016/j.anbehav.2010.07.012)
[11] Zuberbühler K. 2015 Linguistic capacity of non-human animals. *WIREs Cogn. Sci.* **6**, 313-321. (doi:10.1002/wcs.1338)
[12] Radford AN, Ridley AR. 2006 Recruitment Calling: A Novel Form of Extended Parental Care in an Altricial Species. *Curr. Biol.* **16**, 1700-1704. (doi:10.1016/j.cub.2006.06.053)
[13] Engesser S, Ridley AR, Townsend SW. This way! Pied babblers modify the internal structure of their recruitment cries specifying the form of recruitment. Submitted to *Proc. R. Soc. B*.
[14] Raihani NJ, Ridley AR. 2007 Variable fledging age according to group size: trade-offs in a cooperatively breeding bird. *Biol. Lett.* **3**, 624-627. (doi:10.1098/rsbl.2007.0435)
[15] Raihani NJ, Ridley AR. 2007 Adult vocalizations during provisioning: offspring response and postfledging benefits in wild pied babblers. *Anim. Behav.* **74**, 1303-1309. (doi:10.1016/j.anbehav.2007.02.025)
[16] Raihani NJ, Ridley AR. 2008 Experimental evidence for teaching in wild pied babblers. *Anim. Behav.* **75**, 3-11. (doi:10.1016/j.anbehav.2007.07.024)
[17] Bousquet CA, Sumpter DJ, Manser MB. 2011 Moving calls: a vocal mechanism underlying quorum decisions in cohesive groups. *Proc. R. Soc. B* **278**, 1482-1488. (doi:10.1098/rspb.2010.1739)
[18] Sueur C, Deneubourg JL. 2011 Self-Organization in Primates: Understanding the Rules Underlying Collective Movements. *In. J. Primatol.* **32**, 1413-1432. (doi:10.1007/s10764-011-9520-0)

Supplementary material

Methods

Acoustic analyses

To verify the acoustic similarity among cluck and purr elements, we analysed a single cluck and a single purr vocalisation in each of 11 groups. A group-level analysis was favoured as an individual-level analysis would have severely limited our collection of high quality calls given that, due their noisy characteristics and low broadcasting amplitude, both vocalisations are naturally difficult to record. Calls were recorded using a Rode NTG-2 directional microphone coupled with a Rode Blimp Suspension Windshield (Rode microphones) and a Roland R-26 portable recorder (Roland Corporation) (sampling frequency 48kHz, 24-bit accuracy). Only high signal-to-noise ratio vocalisations were selected, and the first element and a further randomly chosen element of each vocalisation were acoustically analysed. Since the elements comprising clucks and purrs are often characterised by atonal, noisy structures, where a clear fundamental frequency (F0) cannot be resolved, we only extracted non-F0 related spectral parameters using Praat (v. 5.3.55) including: element duration, peak frequency, time to maximum intensity, amplitude variation, as well as the first, second and third energy quartiles.

Natural observations

To investigate whether and how clucks influence movement between foraging sites, we collected two sets of observational data: the first set included natural observations conducted at dawn at a group's roost, before the birds started to forage. We recorded the number of individuals producing clucks (including cases when no clucks were produced), the group size, and the distance the group subsequently moved to their first foraging site. Foraging sites were classified as locations where at least half of the group would subsequently forage. Distances were measured using a GPS data logger (eTrex 10, Garmin). The second set of observations was conducted later in the morning or in the early evening when groups were continuously foraging, collecting the same behavioural data as for the first set, but considering only events when clucks were produced, since data on no-clucking events (including the duration of clucking) and subsequent movements are in their nature impossible to collect.

Statistical analyses

Statistical analyses were performed in R (v. 3.2.3) [1]. Model assumptions were inspected visually, and potential multicollinearity among predictor variables was controlled for by removing variables with variance inflation factors (VIF) higher than 5 [2, 3]. To verify the acoustic similarity among the elements within, as well as across clucks and purrs, a cross validated Discriminant Function Analysis (cDFA) was conducted [4, 5]. Due to a high VIF, the frequency measurements at the second energy quartile were excluded from the cDFA. In accordance with the 4 classes to be discriminated, a binomial test with a 25%-change level of

correct classification was conducted to assess the classification probabilities of the cDFA. Linear Mixed Models were fitted to test the effect of the number of clucking individuals, a group's size and a group's overall clucking duration on the distance a group moved (i) from the roost to the first foraging site, and (ii) between two foraging sites [6, 7]. Response variables were transformed if necessary. Since multiple observations per group were used, group identity was fitted as random term. Best models were selected based on the Akaike's information criterion corrected for small sample sizes (AICc), with a threshold difference great than one to the next best model [8]. Parametric bootstrapping was applied to assess the significance of the fixed effects (pbkrtest, 10,000 iterations [9]).

Results

Table S1: Summary of statistical models. Full models included all fixed and random effects, intercept models included only the random effect, and best models included only the significant fixed effects based on AICc selection as well as the random effect. df=degrees of freedom.

Predictor variables	AICc	df
Roosting context		
Full model: clucking duration + group size * number of clucking individuals	218.8	7
Intercept model	232.0	3
Best model: number of clucking individuals	212.9	4
Foraging context		
Full model: clucking duration + group size * number of clucking individuals	603.3	7
Intercept model	609.9	3
Best model: number of clucking individuals	597.1	4

References

[1] R-Core-Team. 2014 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[2] Fox J, Weisberg W. 2011 *An {R} Companion to Applied Regression, Second Edition*. Thousand Oaks CA: Sage.

[3] Frank AF. 2011 R-hacks/mer-utils.R. https://github.com/aufrank/R-hacks/blob/master/mer-utils.R. Accessed 29 Apr 2016.

[4] Venables WN, Ripley BD. 2002 *Modern Applied Statistics with S*. New York: Springer.

[5] Mundry R, Sommer C. 2007 Discriminant function analysis with nonindependent data: consequences and an alternative. *Anim. Behav.* **74**, 965-976. (doi:10.1016/j.anbehav.2006.12.028..)

[6] Barto'n K. 2014 MuMIn: multi-model inference. R package version 1.10.5. http://CRAN.R-project.org/package=MuMIn. Accessed 29 Apr 2016.

[7] Bates D, Maechler M, Bolker B, Walker S. 2014 Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1-48.

[8] Burnham KP, Anderson DR. 2002 *Model selection and multimodel inference: a practical information-theoretic approach*. New York: Springer.

[9] Halekoh U, Højsgaard S. 2014 A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models - TheRPackagepbkrtest. *J. Stat. Softw.* **59**, 1-32. (doi:10.18637/jss.v059.i09)

CHAPTER 6

General discussion

One major question remaining in the field of language evolution is how the generative capacity of human language emerged [1, 2]. This ability, that allows the communication of limitless thoughts, is a result of the combinatorial and compositional nature of human language: meaningless phonemes are combined to create meaningful morphemes or words (phonology), and morphemes or words are assembled into higher-order meaningful structures (syntax/compositionality) [3-5]. Work searching for the evolutionary origin of and selective drivers for combinatorial and compositional abilities has mainly focused on animal songs or the discrete vocalisations of primates [1]. The overall aim of this dissertation was to take a broader comparative approach and investigate the prevalence and forms of combinatorial mechanisms in two passerines that produce discrete (mammal-like) vocalisations: the southern pied babbler (*Turdoides bicolor*) and the chestnut-crowned babbler (*Pomatostomus ruficeps*). As such, this dissertation provides some of the first evidence for rudimentary phonemic and semantically compositional structures outside of humans, and demonstrates further empirical evidence for the prevalence and diversity of combinatoriality in animal communication systems.

Summary of findings

Playback experiments on wild chestnut-crowned babblers in a standardised aviary setup revealed that the birds generate two functionally distinct vocalisations by rearranging two meaningless sounds: the *AB*/'flight' call and the *BAB*/'prompt' call. The meaning-differentiation between the two calls was demonstrated to be the result of a single modification (i.e. the prompt calls first *B* note). While similar forms of sound combination have also been shown in parids (Paridae), the exact function of these parid calls remain ambiguous and rather represent behaviourally specific combinatorial structures [6, 7]. As such, this work on the chestnut-crowned babbler is the first to show that animals are capable of combining acoustically differentiable elements to encode functionally distinct meaning, representing a rudimentary form of phoneme-like structuring, a combinatorial operation omnipresent in human languages.

Work on free-living southern pied babblers unveiled several forms of combinatorial structuring. Firstly, pied babblers were found to produce 'alert' calls in response to sudden, low-urgency threats, and 'recruitment' calls when recruiting group members during group travel. Upon encountering terrestrial predators both vocalisations were combined into a 'mobbing sequence', and playback experiments indicated that receivers responded to the combination in a compositional way. While recruitment calls elicited a slow, steady approach to the simulated caller, receivers quickly approached the 'putative threat' communicated by the combination and were highly alert in response to played back mobbing sequences, with babblers potentially linking the information on the threat and the requested action. The pied babbler mobbing sequence therefore appears to constitute a semantically compositional structure, with the sequence's meaning reflecting the meanings of its compounds. A recent study has revealed a similar call construct in Japanese great tits (*Parus major*

minor), with the tits combining calls that alone elicit scanning and approaching behaviour respectively [8]. However, in contrast to the babbler mobbing sequence, the tit call combination specifically elicits a mixed, intermediate response of the behaviours otherwise elicited by the individual compounds [8]. Accordingly, the compound signal might average-out the information of the single compounds, with the combination potentially better resembling a 'mixed compound call' than a compositional structure [9]. Furthermore, reversing the order of the compounds within the call sequence failed to elicit a response in tits, which has recently been argued to present evidence against a systematic processing of the structure and against compositional semantics [10, 11].

Additional work investigating the acoustic structure of southern pied babbler recruitment calls indicated these vocalisations can also be produced in form of a long 'recruitment cry', with male pied babblers producing two cry variants. Both variants were found to be introduced by a wind-up, grading into a repetition of either *A'*/single-note-motifs, or *A'B'*/double-note-motifs respectively. While both types seem to function to recruit group members, the cries' internal structure (i.e. motif pattern) specified the form of recruitment, with the addition of *B'* notes modifying the recruitment from requesting approach to following over a long distance. Accordingly, recruitment cries appear to be characterised by meaning-differentiating within-structure variation, representing what seems to be a novel type of combinatorial structuring at the internal level of a larger meaningful signal.

Lastly, this dissertation provides evidence for a temporally-based combinatorial mechanism in southern pied babbler 'clucks' and 'purrs'. Both call types were composed of repetitions of the same discrete acoustic element, with the number of repetitions differentiating the two calls. While purrs function to draw dependent offspring to the caller's food patch [12-14], behavioural observations indicated that clucks seem to induce a decision-making process by the group, potentially mediating imminent foraging site switches. Such meaning-differentiating temporal modifications have so far only been demonstrated in the predator specific call sequences of colobus monkeys (*Colobus guereza* & *C. polykomos*) and in behaviour specific calls of Mexican free-tailed bats (*Tadarida brasiliensis*) [15-17]. This additional evidence in pied babblers suggests that temporal structuring may be a common phenomenon in animal communication systems to encode discrete information sets.

Factors promoting combinatoriality

Comparative work on non-human animals can help to identify selective forces promoting the evolution of a particular trait [1]. Specifically, work investigating combinatorial mechanisms in animal communication systems can yield crucial insights into potential precursors and the evolutionary progression of language's combinatorial layers [1]. While studies on the combinatorial abilities in animals commonly assume that the emergence of vocal combinations is the result of small

vocal repertoires paired with the need to encode a variety of messages [8, 18, 19], less is known about the underlying mechanisms and selective conditions driving a species to evolve a certain combinatorial mechanism. The following concluding paragraphs will discuss the implications of our knowledge on animal vocal combinations for the evolution of generative vocal mechanisms, with a focus on the empirical data demonstrated in babblers.

Vocal constraints and signal perception

Most animals are constrained in the number of auditorily distinguishable sounds they can produce and have limited abilities to modify frequency characteristics of their pre-existing vocalisations [20-22]. Changes of temporal characteristics are, on the other hand, thought to be less constrained by the vocal apparatus [20]. Accordingly, it is surprising that temporal mechanisms encoding discrete, categorical information have so far only been demonstrated in few species [15-17]. Perhaps because they do not resemble language's combinatorial layers, temporal aspects have generally been neglected. However, the new insight from pied babbler clucks and purrs suggests that temporal structuring might be, given its combinatorial simplicity, a valuable and widespread mechanism applied by vocally constrained animals to trade-off productional limitations.

A more commonly investigated strategy to increase communicative output is the recombination of (meaningless) sounds to encode diverse information [4, 23]. Theoretical work proposes that, beyond increasing the size of a repertoire, stringing sounds together primarily serves to enhance the discriminability between otherwise similar sounding signals [24, 25]. However, mathematical models appear to be based on more advanced sound rearrangements [24, 25], while the combinatorial mechanisms demonstrated in animals seem to be the result of more simple element omissions and duplications [26]. For example, chestnut-crowned babblers reuse two meaningless elements to create the functionally distinct *AB*/flight and *BAB*/prompt calls, with the meaning-differentiation between the two calls being the result of the absence/presence of a single element (i.e. the prompt call's first *B* note). Similarly, southern pied babblers produce two context specific recruitment cry variants that are only differentiable by their internal structure, either comprising repetitions of *A'*/single-note motifs, or *A'B'*/double-note motifs, with the *B'* notes being the sole difference between the two cry types. Similar mechanisms are also applied by parids, which produce dozens of call variants (although functionally ambiguous) by duplicating or omitting notes within a fixed-ordered sequence [26]. From an evolutionary perspective such 'skipping' mechanisms might be less demanding to produce and develop than potentially more complex element reorganisations. However, the more subtle acoustic changes applied by babblers (and parids [6]) are not without limitations. Specifically, in comparison to element reorganisations (akin to *tin* & *nit*), element skipping (akin to *tin* & *in*) may be more prone to perception mistakes with the risk to confuse signals. Nevertheless, babblers (and parids [6]) seem

to minimise the consequence of misperceived signals, by assigning the vocalisations to non-dangerous/social contexts where mistakes are less likely to incur high costs. Furthermore, chestnut-crowned babbler flight and prompt calls are broadcast over short distances where signal degradation is likely to be minimal. This contrasts with the internal *A'* and *A'B'* motifs of pied babbler recruitment cries which are repeated multiple times generating arguably more easily distinguishable combinations that can be reliably perceived over longer distances. In conclusion, while anatomical constraints might limit the number of distinguishable sounds an animal can produce, sound combinations might have evolved to increase signal discriminability between otherwise physically similar sounds (i.e. *A* & *B* and *A'* & *B'*). The fact that both combinatorial mechanisms applied by chestnut-crowned and southern pied babblers are highly bound (i.e. both babblers generate only two signals out of the two sounds), supports the hypothesis that (at least in babblers) sound combinations primarily emerged to ensure reliable signal perception, ultimately trading-off vocal constraints [24, 25].

Memory limitations and non-redundancy

While sound combinations are suggested to have evolved in order to reduce perception mistakes among otherwise similar sounding signals [24, 25], theory further hypothesises that compositionality (i.e. the combination of meaningful units) emerged to trade-off memory limitations [24, 27]. Specifically, while the association between a signal and its meaning has to be memorised, memory constraints which may limit the number of potential messages that can be communicated, can theoretically be overcome by combining meaningful signals in a systematic way into higher-order structures with a *derived* meaning [24, 27]. However, for such a systematic (or syntactic/compositional) system to be evolutionarily advantageous over the use (and evolution) of discrete signals, the number of messages that can be encoded by a given construct has to exceed the number of discrete signals the construct is composed of [24, 27, 28]. In line with this criteria, the work on southern pied babblers demonstrates by combining alert calls (and potentially aerial and begging calls) with recruitment calls when prompting group members to mob a terrestrial threat, babblers effectively encode 3 concepts by recombining 2 meaningful calls (or potentially 7 concepts by recombining 4 distinct calls respectively). Similar findings have been demonstrated in the affixation system of Campbell's monkeys (*Cercopithecus campbelli campbelli*) with a meaning-modifying element being affixed to either one of two predator specific alarm calls [19, 29], as well as in the concatenated structures of Diana monkeys (*Cercopithecus diana*) and banded mongooses (*Mungos mungo*), with fixed individual-coding segments being recombined with variable behaviour/motivation coding segments [30-32].

The demonstration of such 'sets' of meaningful call combinations (or concatenations) in animals supports the hypothesis of compounds having to be reused in more than one construct for the construct to have a selective advantage over the use

of a novel vocalisation. However, demonstrating evidence for compositional-like structures in animal communication systems might, in fact, contradict the theory that compositionality primarily evolved to counteract memory limitations. In contrast to human languages, which encompass vast lexicons of learned and memorised words [24, 27], animal vocal repertoires differ fundamentally in the size of their lexicons and comprise far less signals [33-35], with memory capacities in animals being unlikely to near their limit. Nevertheless, combining signals in a meaningful, systematic/ compositional way (e.g. alert + recruitment = recruitment to threat) is likely more efficient than learning a new, redundant association between a novel signal and a concept (assuming the cognitive abilities to parse compositional structures are already in place). In fact, it is suggested that a new "cognitive file" is unlikely to be created for an object that is already stored in the animal's memory, as it would create redundancy among the encoded concepts [36]. Thus, compositionality might primarily have evolved to avoid redundancy within a communication system, which in turn might release memory capacities.

Implications for the evolution of human language

The empirical data provided in this dissertation indicates that neither meaningful combinations generated from acoustically differentiable meaningless elements, or meaningful compositional structures, in their basic form, are unique to the human language. The work on chestnut-crowned babbler flight and prompt calls, and southern pied babbler recruitment cries supports the theory that sound combinations primarily emerged to enhance the discriminability between signals, rather than to increase the signal repertoire per se [24, 25]. Furthermore, sound combinations were found to be the result of relatively simple 'skipping' mechanisms, with meaning-differentiation between signals being achieved through the presence or absence of individual sounds. One consequence of these findings is that it can shed important light on the evolutionary progression of language's phonology. For example, such single modifications might have been a potential prerequisite driving sensitivity to initial phoneme structures and the ability to recognise signals as composed of smaller meaningless elements, ultimately promoting the evolution of a full blown phonological system [37]. The demonstration of compositionality in the southern pied babbler mobbing sequence provides further evidence that language's syntactic system can be decomposed into more primitive, 'intermediate' layers, and, as such, opposes the alternative theory postulating that language's syntactic capacity (comprising hierarchical organisations and long-distance dependencies) evolved as an "all-or-nothing package" [38, 39]. Furthermore, empirical evidence for meaningful call combinations in the small vocal repertoires of diverse animals suggests language's compositional nature might have more likely evolved to avoid redundancy among encoded concepts, eventually trading-off memory limitations.

Future work now has to investigate the perceptual and cognitive capacities in the two babbler species. Specifically, in order to further support that stringing

meaningless sounds together creates more easily distinguishable signals, experimental work has to demonstrate that sound constructs (e.g. *XY* & *YXY*) are in fact more reliably perceived and distinguishable than a construct's individual sounds (e.g. *X* & *Y*). Additionally, in order to provide further support for compositional capacities in babblers, the creation of artificial semantically meaningful compositions could determine whether the birds are also able to decompose the meaning of a novel construct from its individual meaningful compounds.

By demonstrating evidence for meaningful vocal combinations in species distantly related to humans, this dissertation provides fundamental insights into the factors that might have kick-started the evolutionary progression of language's generative system. Crucially, the steady accumulation of empirical data in diverse species implies that combinatorial mechanisms are most prevalent in social species [40]. While combinatorial mechanisms are suggested to ultimately enable an increased communicative output, they might be of particular importance in social animals where selection has created the need for specialised signals in order to efficiently coordinate a plethora of activities these systems rely on [41]. In conclusion, language's complex generative design might have arisen from primitive needs to transmit reliably perceivable and non-redundant information. Whilst human language potentially started off with rudimentary element rearrangements and loose signal compounds, such basic combinatorial mechanisms may have become further elaborated, adapting to the communicative needs of humans and facilitating the expression of an ever increasing suite of ideas and novel messages.

References

[1] Hauser MD, Chomsky N, Fitch WT. 2002 The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* **298**, 1569-1579. (doi:10.1126/science.298.5598.1569)

[2] Christiansen MH, Kirby S. 2003 Language evolution: consensus and controversies. *Trends Cogn. Sci.* **7**, 300-307. (doi:10.1016/s1364-6613(03)00136-0)

[3] Hockett CF. 1960 The Origin of Speech. *Sci. Am.* **203**, 88-111.

[4] Marler P. 1977 *The structure of animal communication sounds. Recognition of complex acoustic signals: report of Dahlem workshop*. Berlin: Abakon Verlagsgesellschaft.

[5] Hurford J. 2011 Linguistics from an Evolutionary Point of View. In *Handbook of the Philosophy of Science: Linguistics* (eds. R Kempson, M Fernando & N Ashe), pp. 473-498. Amsterdam: Elsevier Science.

[6] Hailman JP, Ficken MS. 1986 Combinatorial animal communication with computable syntax: Chick-a-dee calling qualifies as 'Language' by structural linguistics. *Anim. Behav.* **34**, 1899-1901. (doi:10.1016/s0003-3472(86)80279-2)

[7] Ficken MS, Hailman ED, Hailman JP. 1994 The Chick-a-Dee Call System of the Mexican Chickadee. *Condor* **96**, 70-82. (doi:10.2307/1369065)

[8] Suzuki TN, Wheatcroft D, Griesser M. 2016 Experimental evidence for compositional syntax in bird calls. *Nat. Commun.* **7**, 10986. (doi:10.1038/ncomms10986)

[9] Cheney DL, Seyfarth RM. 1990 *How mokneys see the world*. Chicago: The University of Chicago Press.

[10] Fujita K. 2016 On the parallel evolution of syntax and lexicon: A Merge-only view. *J. Neurolinguist.* In press. (doi:10.1016/j.jneuroling.2016.05.001)

[11] Phillips S, Wilson WH. 2016 Commentary: Experimental evidence for compositional syntax in bird calls. *Front. Psychol.* **7**, 1171. (doi:10.3389/fpsyg.2016.01171)

[12] Radford AN, Ridley AR. 2006 Recruitment Calling: A Novel Form of Extended Parental Care in an Altricial Species. *Curr. Biol.* **16**, 1700-1704. (doi:10.1016/j.cub.2006.06.053)

[13] Raihani NJ, Ridley AR. 2007 Adult vocalizations during provisioning: offspring response and postfledging benefits in wild pied babblers. *Anim. Behav.* **74**, 1303-1309. (doi:10.1016/j.anbehav.2007.02.025)
[14] Raihani NJ, Ridley AR. 2008 Experimental evidence for teaching in wild pied babblers. *Anim. Behav.* **75**, 3-11. (doi:10.1016/j.anbehav.2007.07.024)
[15] Bohn KM, Schmidt-French B, Ma ST, Pollak GD. 2008 Syllable acoustics, temporal patterns, and call composition vary with behavioral context in Mexican free-tailed bats. *J. Acoust. Soc. Am.* **124**, 1838-1848. (doi:10.1121/1.2953314)
[16] Schel AM, Tranquilli S, Zuberbühler K. 2009 The alarm call system of two species of black-and-white colobus monkeys (*Colobus polykomos* and *Colobus guereza*). *J. Comp. Psychol.* **123**, 136-150. (doi:10.1037/a0014280)
[17] Schel AM, Candiotti A, Zuberbühler K. 2010 Predator-deterring alarm call sequences in Guereza colobus monkeys are meaningful to conspecifics. *Anim. Behav.* **80**, 799-808. (doi:10.1016/j.anbehav.2010.07.012)
[18] Arnold K, Zuberbühler K. 2006 Language evolution: Semantic combinations in primate calls. *Nature* **441**, 303. (doi:10.1038/441303a)
[19] Ouattara K, Lemasson A, Zuberbühler K. 2009 Campbell's Monkeys Use Affixation to Alter Call Meaning. *PLoS ONE* **4**, e7808. (doi:10.1371/journal.pone.0007808)
[20] Janik V, Slater P. 1997 Vocal Learning in Mammals. *Adv. Stud. Behav.* **26**, 59-99. (doi:10.1016/s0065-3454(08)60377-0)
[21] Fitch WT. 2010 *The evolution of language*. Cambridge: Cambridge University Press.
[22] Seyfarth RM, Cheney DL. 2010 Production, usage, and comprehension in animal vocalizations. *Brain. Lang.* **115**, 92-100. (doi:10.1016/j.bandl.2009.10.003)
[23] Yip MJ. 2006 The search for phonology in other species. *Trends Cogn. Sci.* **10**, 442-446. (doi:10.1016/J.Tics.2006.08.001)
[24] Nowak MA, Krakauer DC. 1999 The evolution of language. *Proc. Natl. Acad. Sci. USA* **96**, 8028-8033. (doi:10.1073/pnas.96.14.8028)
[25] Nowak MA, Krakauer DC, Dress A. 1999 An error limit for the evolution of language. *Proc. R. Soc. B* **266**, 2131-2136. (doi:10.1098/rspb.1999.0898)
[26] Hailman JP. 1989 The Organization of Major Vocalizations in the Paridae. *Wilson Bull.* **101**, 305-343.
[27] Nowak MA, Plotkin JB, Jansen VAA. 2000 The evolution of syntactic communication. *Nature* **404**, 495-498. (doi:10.1038/35006635)
[28] Nowak MA, Komarova NL. 2001 Towards an evolutionary theory of language. *Trends Cogn. Sci.* **5**, 288-295. (doi:10.1016/S1364-6613(00)01683-1)
[29] Coye C, Ouattara K, Zuberbühler K, Lemasson A. 2015 Suffixation influences receivers' behaviour in non-human primates. *Proc. R. Soc. B* **282**, 20150265. (doi:10.1098/rspb.2015.0265)
[30] Candiotti A, Zuberbühler K, Lemasson A. 2012 Context-related call combinations in female Diana monkeys. *Anim. Cogn.* **15**, 327-339. (doi:10.1007/s10071-011-0456-8)
[31] Jansen ADWAM. 2013 Vocal Communication in Banded Mongoose, University of Zurich, PhD.
[32] Coye C, Zuberbühler K, Lemasson A. 2016 Morphologically structured vocalizations in female Diana monkeys. *Anim. Behav.* **115**, 97-105. (doi:10.1016/j.anbehav.2016.03.010)
[33] Pollard KA, Blumstein DT. 2012 Evolving communicative complexity: insights from rodents and beyond. *Philos. T. R. Soc. B* **367**, 1869-1878. (doi:10.1098/rstb.2011.0221)
[34] Manser MB, Jansen DAWAM, Graw B, Hollén LI, Bousquet CAH, Furrer RD, le Roux A. 2014 Vocal Complexity in Meerkats and Other Mongoose Species. *Adv. Stud. Behav.* **46**, 281-310. (doi:10.1016/b978-0-12-800286-5.00006-7)
[35] Crane JMS, Savage JL, Russell AF. 2016 Diversity and function of vocalisations in the cooperatively breeding Chestnut-crowned Babbler. *Emu* **116**, 241. (doi:10.1071/mu15048)
[36] Schlenker P. 2005 Non-Redundancy: Towards a Semantic Reinterpretation of Binding Theory. *Nat. Lang. Seman.* **13**, 1-92. (doi:10.1007/s11050-004-2440-1)
[37] Anthony JL, Francis DJ. 2005 Development of Phonological Awareness. *Curr. Dir. Psychol. Sci.* **14**, 255-259. (doi:10.1111/j.0963-7214.2005.00376.x)
[38] Berwick RC. 2011 All you need is merge: biology, computation, and language from the bottom-up. In *The Biolinguistic Enterprise* (eds. AM Di Sciullo & C Boeckx), pp. 706–825. Cambridge, MA: The MIT Press.
[39] Progovac L. 2015 *Evolutionary Syntax*. Oxford: Oxford University Press.
[40] Collier K, Bickel B, van Schaik CP, Manser MB, Townsend SW. 2014 Language evolution: syntax before phonology? *Proc. R. Soc. B* **281**, 20140263. (doi:10.1098/rspb.2014.0263)
[41] Hauser MD. 1997 *The Evolution of Communication*. Cambridge: MIT Press.

CURRICULUM VITAE

PERSONAL DETAILS

Name:	ENGESSER
First Name:	Sabrina
Date of Birth:	1th of May 1984
Nationality:	German

EDUCATION AND QUALIFICATIONS

2012 – 2016	**PhD**, IEU, University of Zurich, Switzerland *Thesis: "Vocal combinations in the southern pied babbler (Turdoides bicolor) and the chestnut-crowned babbler (Pomatostomus ruficeps): implications for the evolution of human language", supervision Dr. Simon W. Townsend, Prof. Marta B. Manser & Dr. Amanda R. Ridley*
2009 – 2011	**MSc in Biology - Behavioural Sciences,** IEU, University of Zurich, Switzerland *Thesis: "The function of close calls in a group foraging carnivore, Suricata suricatta", supervision Prof. Marta B. Manser*
2006 – 2009	**BSc in Biology**, TU Kaiserslautern, Germany *Thesis: "Wild bee diversity on the white clover in different habitat structures of the Palatinate Forest", supervision Dr. Jürgen Kusch*
2004 – 2006	**A-level**, Technical College Singen, Germany
2000 – 2004	**GNVQ-level**, Vocational Secondary School Rottweil, Germany
1994 – 2000	**O-level**, Secondary Modern School Immendingen, Germany

PROFESSIONAL EXPERIENCE

2011 – 2012	**Research assistant**, Prof. Marta B. Manser, IEU, University of Zurich, Switzerland
2000 – 2004	**Apprenticeship** & **Employment as Technical Drawer**, Chiron-Werke GmbH & Co. KG, Tuttlingen, Germany

GRANTS & AWARDS

2016	**Material prize** for top-5 student oral presentations at the Evolang 2016 conference
2015	**Conference travel grant**, IEU University of Zurich, 550.90 CHF
2014 – 2015	**PhD scholarship**, Forschungskredit University of Zurich (FK-14-077), 60'617 CHF
2014 – 2014	**Mobility fellowship**, SNF (P1ZHP3_151648), 35'967 CHF
2014	**Conference travel grant**, IEU University of Zurich, 2'399 CHF
2012 – 2013	**PhD scholarship**, Forschungskredit University of Zurich (57191601), 67'000 CHF

PUBLICATIONS

- Engesser S, Crane JMS, Savage JL, Russell AF, Townsend SW. (2015) Experimental evidence for phonemic contrasts in a nonhuman vocal system. *PLoS Biol.* 13(6): e1002171
- Engesser S, Ridley AR, Townsend SW. (2016) Meaningful call combinations and compositional processing in the southern pied babbler. *PNAS* 24(113):5976-5981
- Engesser S, Ridley AR, Manser MB, Townsend SW. This way! Pied babblers modify the internal structure of their recruitment cries specifying the form of recruitment. *Submitted to Proc. R. Soc. B.*
- Engesser S, Ridley AR, Townsend SW. Element repetition rates encode functionally distinct information in pied babbler 'clucks' and 'purrs'. *To be submitted*

PHD PROGRAM IN EVOLUTIONARY BIOLOGY

TRANSFERABLE SKILLS

BIO338 Course in Scientific Writing	2 ECTS
UWW275 Introduction to Programming in R	1 ECTS
Führungskompetenzen für Doktorierende	1 ECTS

EVOLUTIONARY BIOLOGY

BIO554 Topics in Evolutionary Biology	1 ECTS

OTHERS

BIO215 Evolution of Human Nature	2 ECTS
123C Rätsel um die gesprochen Sprache	4 ECTS
BIO620 Literature Seminar Animal Behaviour	1 ECTS

SCIENTIFIC COMMUNICATIONS & WORKSHOPS

11/2016 **Invited seminar**, Dagstuhl Seminar 16442, Schloss Dagstuhl, Germany: *Vocal Interactivity in-and-between Humans, Animals and Robots (VIHAR)*

10/2016 **Invited talk**, Max Planck Institute for Ornithology Seewiesen, Germany: oral presentation "*Vocal combinations in the southern pied babbler and the chestnut-crowned babbler: implications for the evolution of human language*"

07/2016 **Cambridge Meeting**, University of Cambridge, UK: oral presentation *"Compositional syntax in pied babblers"*

06/2016 **Workshop**, University of Zurich, Switzerland: *Animal communication: signal complexity - what questions to ask?*

04/2016 **Evolang 2016**, New Orleans, USA: oral presentation "*Meaningful call combinations and compositional processing in a social bird*"

12/2015 **Behaviour, Ecology, Environment and Evolution Seminar, IEU University of Zurich**, Switzerland: oral presentation "*Combinatoriality in two social birds*"

09/2015 **Protolang 2015**, Rome, Italy: oral presentations "*Call combinations and compositional processing in a social bird*", and on behalf of S. Townsend "*Experimental evidence for phonemic contrasts in a non-human vocal system*"

08/2014 **ISBE 2014**, New York, USA: oral presentation "*A precursor of phonemic contrasts in a non-human vocal system*"

05/2014 **Workshop**, University of Neuchatel, Switzerland: *Methods in animal communication research - calls, facial expressions, gestures*

06/2013 **Annual Research Symposium of the PhD program**, Möschberg, Switzerland: *Evolutionary topics in biology*

12/2012 **ProDoc workshop on proximate and ultimate causes of cooperation**, Beatenberg, Switzerland: oral presentation "*Call combinations in the pied babbler*"

10/2012 **PhD CogBio Vienna visit in Zurich**, University of Zurich, Switzerland: oral presentation "*Call combinations in the pied babbler*"

Babblers are fluffy and awesome!